THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXXIV April, 1943 Number 4

RATE OF READING AND LEARNING BY READING

J. B. STROUD AND MARGARET HENDERSON*
State University of Iowa

The claim that fast readers are better than slow readers rests upon data that are not entirely adequate. If fast readers are better than slow readers, this may be taken to mean they 'get more' out of what they read, learn more about the content of the reading matter. It is doubtful whether or not any general statement can be made about the relationship between rate and comprehension. Much would seem to depend upon the conditions of reading and the methods of measurement. Moreover, since reading is a many-sided activity, it cannot be said that any one method of measurement is superior in all respects to the others. There are several approaches to the problem, each of which is valid within a given set of operations.

The most widely used procedure has been that of correlating rate scores with comprehension scores earned on standardized reading tests. This procedure has yielded coefficients, characteristically, of from .40 to .50. These findings led to the conclusion that quality of reading is positively correlated with rate. The data do not warrant this generalization. The comprehension score—the number of items answered correctly, with perhaps a correction for guessing—is to some unknown degree a rate score, since the reader works against time. The number he gets right depends to some extent on how fast he can read the text material and the questions.

In the foregoing procedure the rate score is determined on one selection of reading matter and the comprehension score on another selection, the two operations being independent of each

* The first two experiments reported in this article were conducted by the writer, the first-named author. The third experiment was conducted by Henderson and represents a portion of the work done for her Master's thesis.

other. By another procedure, rate and comprehension scores may be derived from the same operation, the number of items attempted constituting the rate score and the number of items responded to correctly constituting the comprehension score. This procedure would doubtless give a high correlation between the two scores if the test were reliable. If the test were primarily a speed test, the two scores would tend to be very largely the same scores. These two scores would probably also intercorrelate significantly in power-test situations, for the reason that power makes for accuracy and, on a power test, also makes for speed. In a test situation in which an attempt is made to control the quality of reading by the employment of standard questions and problems, rate is determined by 'thinking time' as well as by reading time. The able student certainly has an advantage in thinking time, in a power situation. The only objection to a procedure that forces thinking time to become a large element in reading rate is that it is unnatural. The reader is not normally confronted with lists of questions to which right answers are then and there called for. We shall return to this question later.

Flanagan obtained a correlation of .17 between rate of reading, as determined by the last item attempted in standard time, and 'level of comprehension,' in still another approach to the problem.[3,*] This approach is not to be confused with that referred to in the preceding paragraph. Flanagan's procedure eliminates the speed factor in the comprehension score, and thus overcomes the objection to the first-mentioned procedure. The obtained coefficient signifies the presence of comparatively little relationship between the two measures. The procedure seems to be more adequate for determining the relationship between rate and comprehension than any procedure heretofore employed, although admittedly one cannot say precisely what 'rate' refers to. It would be possible for a reader to look at the words and pronounce them subvocally without understanding the material. We have

* In a sense 'level of comprehension' is an accuracy or quality score. The test on which Flanagan[3] derived this score was divided into four successive and equal parts of twenty items each. The level of comprehension score is the average number of items answered correctly on the parts completed. A pupil who completes one part and answers fifteen items correctly would earn the same level of comprehension score as one who completes all four parts and maintains an average of fifteen items answered correctly. Thus the maximum level of comprehension score is 20.

to rely upon the subject to make an effort to understand what he reads; in other words, to coöperate—as we do in most experiments in the field of human behavior. Sentences that are not understood, or that are poorly understood, require at least as much time as those that are understood. We cannot, therefore, reckon rate by considering only those sentences that are understood, unless we have suitable means of computing time separately for those that are understood and those that are not.

One of the interpretations placed upon the positive correlations obtained between rate and comprehension (by the conventional method) is that fast readers learn more, understand more, per reading, than do slow readers. If this is the construction placed upon the findings, it would seem feasible to attack the problem directly; that is, to ascertain the relationship between reading rate and learning scores. This method was used several years ago by Abell[1] and later by King.[5] Most of those writers who have taken any account of the method at all have summarily dismissed it on the ground that it yields a 'memory' score. Since the problem is to ascertain the relationship between rate and comprehension, knowledge of the relationship between rate and memory is naturally not germane, so the reasoning seems to go.

It appears that writers and investigators may have been semantically tricked. Psychologically, memory and comprehension are not so dichotomous as the terms might suggest, nor are memory and learning. Presumably it would be possible to construct a memory test that would permit of no admixture of understanding, but this would probably have to be a test of rote memory. As Henderson[4] and Kuhlmann[6] some years ago, and Woodworth[8] and Bartlett[2] more recently, pointed out, remembering is in large measure a process of 'reconstruction.' This is so to a degree even in verbatim memory. One attempts to reconstruct a logical, meaningful whole. Let the reader reread and try to repeat the last three sentences. He will observe that his rendition makes sense. If he cannot remember some of the exact words, he supplies others that carry the same import. He does not leave gaps where the exact words of the text cannot be reproduced, or substitute inappropriate words. Meaningful sentences are reconstructed, as may be said, out of his understanding. In any learning situation that involves understanding, that understanding can be measured a moment later. Those elements in a text, such as dates and names, that do not

lend themselves well to the testing of understanding when employed as items in a test of learning, do not lend themselves well to the testing of the understanding when employed in the conventional reading test. The question reduces to this: is there any sense in distinguishing between a good reader and a good learner when reading is made the medium of learning?

It is possible to prepare an objective test over suitable text material that will measure the understandings, concepts, and meanings gained from reading this text, when administered without a time limitation immediately after the completion of the reading. In fact, one might make a defense of this procedure as a valid method of testing reading proficiency. It is certainly not wanting in naturalness. When a student reads a book, a column in the newspaper, or studies his lessons, he is not supplied with a list of questions to the answering of which his attenion is directed as he reads. He ordinarily reads along in a fairly continuous fashion, giving varying degrees of 'potency' to the words and phrases in accordance with his own interests and previous experience. Likewise, in school and in life we ordinarily appraise a person's understanding of what he reads by what he is able to say about it afterward—how intelligently he can discuss it—and by what use he is able to make of the content in reacting to problems, verbal or otherwise, to which it is pertinent.

In the investigations reported in this article the method described in the foregoing paragraph has been used. While for reasons just stated it is regarded as a feasible method of attacking the problem of the relationship between rate and comprehension, it may be defended on other ground if the reader dissents. It is pertinent to the problem of reading. It is worth knowing what the relationship between rate of reading and learning is. Educationally, this is a vital problem. We cannot assess the extent of the slow reader's handicap, nor the fast reader's advantage, in getting an education in the absence of information on this point. The obtained relationship between rate and comprehension scores on standardized reading tests has been construed as supplying this information. For reasons stated above, such a construction is open to question.

EXPERIMENT I

Two printed selections—*Pellagra* and *John Marshall,* consisting of six hundred forty-two and five hundred eighty-six words,

respectively—taken from the reading section of one of the Iowa Every Pupil Tests of Basic Skills were used as the text material in Experiment I. Since the pupils were to read a selection at their own rate and then respond to the test questions without further reference to the text, some of the questions used in the original test had to be modified or supplanted by other questions. Each test was comprised of fifteen four-response items. The following is a representative item. "What was the difference between the meat Dr. Goldberger fed his patients and the meat they had been eating? (1) He used meat from animals that had no disease; (2) He used beef instead of pork; (3) He used meat that had been properly cured; (4) He used meat that contained little fat." Some of the other questions, for which suitable responses were framed, are as follows: "Which is the best title for this article?" "In addition to the fact that doctors, nurses, and people who received the pellagra blood did not get the disease, what other proof is there that pellagra is not catching?" "What was the first main step in Dr. Goldberger's study of pellagra?" The foregoing will serve to give the reader some idea of the nature of the tests.

Two hundred eighty-eight students, Grades V to VIII, served as subjects in this experiment. These represented the entire enrolment in the grades in question in a certain small city, except for absentees and a few students who failed to record one or both time scores. The students were instructed to read the material "carefully but fairly fast." "Read well and carefully, but do not reread or loaf along. Do not read the article more than once." The experiment was conducted in regular classroom situations. The teachers in charge of the classes assisted with its administration. Time was written on the board every fifteen seconds. At the conclusion of the reading the pupils recorded the figures then appearing on the board. They read the first selection, recorded the time, laid the material aside and proceeded to answer the questions. When all had finished the test, the second selection was read and the second test subsequently taken. No time limits were placed on the tests.

The results of the two tests were combined. The sums of the two time scores were correlated with the sums of the two learning scores. The coefficients for the four grades, in order from the fifth to the eighth, are: .06, .02, .12, and .02.

These coefficients suggest that, under these conditions, reading time and learning are unrelated. It is true, of course, that lack of reliability of scores would lead to the same result. A correlation of .72 was obtained between the two time scores; and a correlation of .56, between the two learning scores. Since the two time scores and also the two learning scores were combined, we may assume that the scores are somewhat more reliable than these coefficients indicate. The mean time score, all grades, is 7′ 8″, the SD being 2′ 56″. The total number of words equals 1228.

Attention is called to a question of procedure. When the time-limit method is used, *i.e.*, when time is kept constant and amount read allowed to vary, the arithmetic mean of words-per-minute scores is a legitimate measure of central tendency. However, when the work-limit method is used, *i.e.*, when amount read is constant and time becomes the variable, as in the experiments reported in this article, the AM is inappropriate. The harmonic mean should be used; or the mean time should be computed and this value converted into mean words per minute. For example, when A reads eight hundred words in two minutes, and B reads eight hundred words in four minutes, A's rate is four hundred words per minute and B's, two hundred. The AM is three hundred words. However, the mean reading time is three minutes; and 800 words ÷ 3 minutes = 266 words per minute.

In experiment I the mean reading is 7′ 8″. The number of words read (1228) divided by 7′ 8″ yields a mean rate score of one hundred seventy-two words per minute. The HM of words-per-minute scores is likewise one hundred seventy-two; however, the AM of these scores is one hundred ninety-six.

EXPERIMENT II

In an investigation of this kind the problem of obtaining representative rate scores is troublesome. There is, of course, no one fixed rate at which a person reads; and the conditions of the experiment are likely to influence rate in a particular reading situation. However, it is probably true that the person who reads fast or slow under one condition reads proportionately fast or slow under other conditions. For Experiment II special exposure apparatus was designed and used. This consisted of a movable half-belt screen, with an aperture sufficient to expose

three lines at a time. The screen was driven by a friction clutch activated by a ten-inch disc driven by a synchronous motor. By adjusting the position of the clutch on the disc the speed of the screen was varied to suit the subject. The reading matter placed under the screen was stationary. Motor noise was perceptible, though not disturbing.

The subjects, taken individually, were asked to read about a page and a half from Shaffer's *Psychology of Adjustment* at the outset, reading at their normal rate. The experimenter took the time with a stop watch. He then set the clutch for the rate at which this selection was read. The apparatus was so calibrated as to make this adjustment fairly accurate, although reading time in the experiment was always taken with a stop watch. The subject was then shown how to make the screen move faster or slower by turning a wheel, and requested to make such adjustments as required during preliminary trials. He was urged to take all the time necessary to adjust the screen to his "normal rate of reading, the rate at which he liked to read."

Three selections of reading matter, each followed by a test, were administered to each subject at a single sitting. When the subject signified that the screen was properly adjusted as to rate, the first selection was presented at that rate. After he had responded to the test questions for this selection, he was informed that he could effect further adjustment in the screen if the rate was not right for him. Four of the ninety-six subjects made slight alterations, in each instance in favor of slower speed. The other two selections were presented at the then existing rate. As in Experiment I, an appropriate test was administered immediately after the reading of each selection. The tests consisted of fifteen, twenty, and twenty four-response items, respectively.

The screen was mounted on a reading desk before which the subject sat. The reading matter was printed in twelve-point type and a seven-inch line, each selection on a single page. The longest page was twelve and one-half inches; the shortest, eleven inches. The selections were 680, 685, and 797 words in length. The three-line exposure permitted the subject to make regressions or otherwise to straighten out a sentence if he got stuck.

The subjects were students of education, nineteen undergraduate and seventy-seven graduate. All were summer-school students; but the graduate students were, on the whole,

considerably older than the undergraduates. The writer was well acquainted with these subjects and feels that the range of talent was at least as great as that found in a typical cross-section of summer-school students of education on a university campus. For all subjects the mean reading time, total for the three selections, was 9′ 57″; with an SD 2′ 12″. This is equivalent to a mean of two hundred seventeen words per minute. The SD is about fifty words per minute. The mean learning test score was 32.34; with an SD of 6.11. The correlation between the combined reading time and the combined learning scores was found to be −.16; which means a correlation of +.16 between rate in words per minute and learning scores.

Incidentally, a correlation of −.37 was obtained between rate —words per minute—and age. The age range was from twenty to fifty-one years. (M = 35; SD = 8.) This finding is at best only suggestive. Even if the size of the sampling were ample, it is not known whether it is equally representative at all age levels. Particularly, the age groups may differ in other respects than age. Learning scores were found to be unrelated to age.

EXPERIMENT III

The third experiment resembles Experiment I as to procedure and materials. The subjects were six hundred twenty-five fifth-grade pupils, representing the fifth-grade population of three schools in Cicero, Illinois, and the entire fifth grade of Bloomington, Highland Park, and six small towns and rural communities, in the same state. The text material consisted of four selections of about twelve hundred words each, two of which were taken from widely used fifth-grade social science books and one each from second- and ninth-grade books in the same field. This putative range in difficulty was provided for a purpose to be mentioned presently. Following the reading of each selection the pupils were given a fifteen-item, four-response test appropriate to the selection. As in the tests used in Experiments I and II, an effort was made to test the pupils' understanding of the materials read.* The materials were administered at two sittings, two selections per sitting, on consecutive days. The reading matter was printed in double columns.

* Prof. Herbert Spitzer, who has constructed the reading tests in certain forms of the Iowa Every Pupil Tests of Basic Skills, supervised the construction of these tests.

For purposes of determining reading time a series of cards, on which were printed random numbers sufficiently large for all to see, was arranged on a rack. A card was turned every ten seconds, after the first minute. These numbers were translated into time scores by means of a prearranged key. Printed instructions to the pupils were, in part, as follows:

> This is an interesting story. It is two pages long. Read it just as you always read your lessons. Read carefully. Do not waste time and do not read the story but once.
>
> Just as soon as you have read both pages, write at the bottom of the page the number on the card in front of you.
>
> As soon as you have written this number at the bottom of the last page, close the booklet, put it on the side of your desk, and do not look at it again.
>
> Take the other booklet you have on your desk and turn it over. Write your name at the top of the page and answer the questions. Choose the answer that best answers the questions on the material you have just read. Work carefully as you will have as much time as you need.

The foregoing instructions were used for the last three of the selections only. The students were given no intimation that a test was to follow the reading of the first selection, one of the fifth-grade selections. The test booklets for this selection were not distributed until after the selections had been read, and no mention was made of a test in the instructions.

The correlations obtained between reading time and the learning scores, for the four selections, are as follows: +.03, −.02, −.05 and +.05. The obtained coefficients between learning scores and reading time, under the conditions of this experiment, suggest a zero relationship between the two variables. These results, as well as those of the two foregoing experiments, give no support to the claim that pupils who read at a rapid rate learn more per reading than those who read at a slow rate. Neither was there any evidence that they learned less, which still makes fast reading altogether desirable.

The reading selections and conditions of this experiment are not strictly comparable, since three levels of difficulty are represented, and since of the two fifth-grade selections, which are presumed to be somewhat equal in difficulty, one was read without knowledge that a test was to follow while such knowledge was supplied for the other. However, some of the inter-correlations

are here listed. The coefficient between the time scores on the second- and ninth-grade selections is .85; that between the two fifth-grade selections, .63. Correlations between the learning scores of the same selections are .49 and .53, respectively.

Since the population represented in this experiment is fairly large and probably fairly representative of the fifth-grade, it may be of some interest to know what the distribution of rate scores, in words read per minute, is on each of the selections. These are, accordingly, given in Table I. The reader will recall that the first fifth-grade selection was read without knowledge that a test was to follow, although some unknown number of subjects may have suspected there would be a test. The pupils were informed in advance that a test would follow the reading of each of the other three selections. The means of the learning scores for the four selections, in the order given in Table I, are 7.59, 7.60, 11.99, and 6.72. The learning scores corresponding to the 90th percentile, in the same order, are 12.18, 11.62, 14.94, and 11.11; and to the 10th percentile, 4.16, 4.60, 8.84, and 3.58.

TABLE I.—SCORES, IN WORDS READ PER MINUTE, BY FIFTH-GRADE PUPILS, CORRESPONDING TO CERTAIN PERCENTILES

Percentiles	Selections			
	Fifth-grade	Fifth-grade	Second-grade	Ninth-grade
99	539	491	794	554
90	338	320	379	361
75	247	230	272	240
50	185	181	206	185
25	149	133	154	143
10	120	105	123	111
1	80	63	69	63

An incidental purpose of this experiment was to secure some information about adaptation of rate to difficulty of selection. No evidence was found that, under these conditions, fast readers are any more adaptable than slow readers, nor any less so. The differences in reading time as between the second- and ninth-grade selections on the part of the fastest ten per cent (as deter-

mined by time scores achieved on the second fifth-grade selection), were approximately equal to the differences exhibited by the slowest ten per cent on these same two selections. However, good learners—and, by one definition, good readers—did show a greater difference in reading time as between the second- and ninth-grade selections than did the poor learners. The best ten per cent (M fifth-grade selection = 12.99) read the second-grade selection at an average rate (AM) of two hundred sixty-five words per minute, and the ninth-grade selection at the rate of one hundred ninety-nine words per minute. The poorest ten per cent (M fifth-grade selection = 3.14) read the second-grade selection at an average rate of two hundred thirty-four words per minute, and the ninth-grade selection at the rate of two hundred forty-nine words per minute. Thus, while there is no evidence, in the results of this experiment, that fast readers are any more adaptable than slow readers, there is some evidence that good learners are more successful in adapting rate to difficulty than are poor learners.

DISCUSSION

It should be emphasized that generalization beyond the conditions of these experiments is not warranted. There probably are valid conditions of reading and methods of measurement that would yield significant positive correlations between rate of reading and understanding. This is a reasonable inference because of the presence of the common factor of intelligence. For example, on a power test in which reading time is defined as the time required to read a selection and solve a series of pertinent reading problems—that is, in which reading time and thinking time are both involved in the rate score—there probably would be found a positive correlation between rate and understanding, as suggested earlier. This should occur because power enables the reader to size up the problem quickly and to arrive at the answer quickly; and this same power would make for a high percentage of accuracy. While speed on a speed test may be unrelated to power, speed on a power test and power are related. On such a test power is conducive to speed.

There should be no objection to the method of measurement involved in the example just referred to. It is no less a valid measure of rate of reading because thinking time becomes a part of the reading-time score. But the results obtained are not the

same as those obtained under other conditions. There are various valid conditions of reading and various valid approaches to the question of the relationship between rate and quality in reading. An answer arrived at by one approach does not immediately become a general answer that applies to all aspects of the problem.

It seems likely that on any power test, which is actually used as a power test—which means that pupils will be asked to answer increasingly difficult questions as they read—speed of reading should correlate positively with the number of questions answered correctly, when the same selection and the same act of reading are made to provide the two scores. Under such an experimental condition power makes for speed and for accuracy. However, in more normal reading situations the power-test situation is present to a smaller degree. The greater the power, the greater is the likelihood that questions and problems will be raised, which will tend to retard rate of reading somewhat. Such questions, were they raised, would retard the reader with limited power even more, but perhaps they are not raised so frequently, or, if raised, not so searching. It is reasonable to suggest that the raising and answering of questions, the sensing and solving of problems, constitute the essence of good reading. But they do tend to make for slower reading. Whereas, the less proficient reader, although probably often slower, inherently, in rate, is not 'hampered' by his own questions, discernments, and reflections to so great a degree. Thus, while under somewhat comparable reading conditions—in which the reading of all is directed to the answering of a common set of questions or solving a common set of problems—the best readers, in a power test, would be expected to proceed at the fastest rate, it does not follow that in natural reading situations this is the case, even on a difficult selection. In Experiment III of this article the correlation between reading time and learning scores was no greater on the ninth-grade selection than on the second-grade selection.

SUMMARY

It has been suggested that the obtained correlations between rate and comprehension scores earned on standardized tests are restricted in meaning. In the first place, the comprehension score is in a measure a rate score, since the reader works against time. In the second place, the obtained relationship may be produced, to some extent, by certain conditions created by the

experimental procedure. These conditions are valid reading conditions, but are not normal or typical, since by the interjection of questions and problems an attempt is made to control the quality or level of reading. Again, this is not objectionable. However, we are not warranted in generalizing from these controlled conditions to natural conditions—not merely because they are controlled, but because they are different. Particularly, we are not warranted, *ipso facto*, in assuming that in life situations, where no attempt is made, experimentally, to control quality or level, there is a relationship of the same kind or of the same magnitude between rate and quality in reading.

The use of learning scores as a measure of reading proficiency has been defended: first, on the ground that it eliminates the time element in the comprehension scores as obtained by the conventional procedure; and, second, on the ground that the reading situation presented is more natural, more typical, and likely to give us a truer picture of the relationship between rate and quality in reading. The results of the investigations reported in this paper suggest, for these conditions, that rate and quality are virtually unrelated. Finally, it is suggested that no general statement can be made about the relationship, other than that it varies with the conditions of reading imposed and with the method of measurement employed.

BIBLIOGRAPHY

1. A. Abell: "Rapid reading: advantages and methods." *Educational review*, Vol. VIII, 1894, pp. 283–286.

2. F. C. Bartlett: *Remembering: a study in experimental and social psychology*. Cambridge: University Press, 1932.

3. J. C. Flanagan: "A new type of reading test for the secondary-school and college students which provides separate scores for speed of comprehension and level of comprehension." *Official report, American educational research association*, March, 1938, pp. 195–199.

4. E. N. Henderson: "A study of memory for connected trains of thought." *Psychological monographs*, Vol. V, 1903, whole no. 23.

5. I. King: "A comparison of slow and rapid readers." *School and society*, Vol. IV, 1916, pp. 830–834.

6. F. Kuhlmann: "On the analysis of the memory consciousness: a study in the mental imagery and memory of meaningless visual forms." *Psychological review*, Vol. XIII, 1906, pp. 316–348.

7. J. B. Stroud: "A critical note on reading." *Psychological bulletin*, Vol. XXXIX, 1942, pp. 173–178.

8. R. S. Woodworth: *Experimental psychology*. New York: Henry Holt and Co., 1938.

ATTITUDE DIFFERENCES BETWEEN AMISH AND NON-AMISH CHILDREN ATTENDING THE SAME SCHOOLS

T. L. ENGLE

Indiana University, Fort Wayne Center

AND

ELEANOR ENGLE

Fort Wayne, Indiana

In his classes in Mental Hygiene one of the writers frequently has teachers who teach in schools attended by children belonging to the Amish sect as well as those belonging to other religious groups. These teachers have often expressed the opinion that because of the unusual customs, including the wearing of a garb, most Amish children are quite maladjusted, that they tend to feel themselves to be "peculiar," that they are excessively submissive and introverted. Certainly to one not belonging to the sect it might well seem that the customs of the Amish would produce many conflicts and unusual attitudes. However, the psychologist may well turn to the individuals themselves in order to measure the effect of such unusual customs upon the personality pattern. The present study has been an attempt to measure certain aspects of personality by means of an inventory presented to Amish school children and to compare the results with those obtained by means of the same inventory presented to non-Amish children attending the same schools. The inventory was administered to the children in the spring of 1942.

For those not familiar with the Amish sect it should be said that the girls wear long, simple dresses of plain color, usually blue, brown, or gray, with no jewelry or other adornments. An apron is worn. Indoors they wear small white or black caps over their hair and upon going outside they put on black bonnets and shawls. The boys wear broad, stiff-brimmed black hats, long dark trousers, and dark shirts with no neckties. The hair cut of the boys is distinctive, the hair usually being worn longer than it is worn by non-Amish boys and often giving the appearance of bangs in front. Both sexes fasten the clothing with hooks and eyes or safety pins rather than with buttons.

Through the years the Amish have driven horses and buggies instead of automobiles; they will not have telephones or radios in their homes; they do not use electric lights; and in other ways they adhere to old practices. The parents, especially the father, are reported to be very domineering in their disciplinary relationships with their children.

In the present study the term Amish is used simply to refer to children wearing the distinctive dress, and the term non-Amish is used to refer to all children not wearing the garb. Some children designated as non-Amish are actually members of a branch of the Amish sect which does not prescribe a garb and which does permit the use of automobiles, telephones, and so forth.

The subjects were children in grades V through VIII in three rural schools in northeastern Indiana. The schools selected were chosen because roughly half of the children were Amish and half were non-Amish in each case. In one of the schools there were forty-three Amish children and thirty-six non-Amish in the grades indicated. In another school the corresponding enrolments were twenty-six and thirty-two; for the third school the enrolments were eleven and eighteen, respectively. The totals for the three schools are eighty Amish children (forty-two boys, thirty-eight girls) and eighty-six non-Amish children (forty-eight boys, thirty-eight girls). Distributions by sex and grades are indicated in Table I.

The personality inventory used was the Pintner-Loftus-Forlando-Alster *Aspects of Personality* (World Book Co., 1937). This inventory measures three aspects of personality: ascendance-submission, extraversion-introversion, and emotional stability. It is definitely designed for use with pupils at the elementary-school level. Two questions refer to behavior on street-cars, but many of the children in the present study had had no personal experience with street-cars so they were told to substitute "bus" or "school bus" for the word "street-car" in those questions. All questions on the inventory are to be answered by crossing out a small square containing either an *S* or a *D*; crossing out the *S* indicates that the child feels the same as the given statement; crossing out the *D* indicates that the child feels different from the statement.

TABLE I.—DISTRIBUTION OF ASCENDANCE-SUBMISSION SCORES BY SEX, SCHOOL GRADE, AND RELIGIOUS SECT

Sex	School Grade	Sect	Number of Subjects	Mean Score	SE Sample	SE Mean	Diff. in Means	SE Diff.	D/SE Diff.
Boys	V–VI	Amish	21	13.62	4.32	0.97	1.19	1.41	0.84
		Non-Amish	21	14.81	4.54	1.02			
	VII–VIII	Amish	21	10.71	4.02	0.90	2.66	1.13	2.35
		Non-Amish	27	13.37	3.50	0.69			
Girls	V–VI	Amish	23	12.96	4.88	1.04	3.44	1.60	2.15
		Non-Amish	20	16.40	5.28	1.21			
	VII–VIII	Amish	15	11.40	2.85	0.76	1.66	1.27	1.31
		Non-Amish	18	13.06	4.22	1.02			
Both	V–VIII	Amish	80	12.25	4.35	0.49	2.11	0.69	3.06
		Non-Amish	86	14.36	4.55	0.49			

All children in grades V through VIII in the three schools under study were given the personality inventory. Mean scores for various Amish and non-Amish groups were computed and the significances of the differences were calculated as indicated in Tables I, II, and III. Means for boys and girls were computed separately. Also, children in the fifth and sixth grades were compared with children in the seventh and eighth grades. Such comparisons resulted in small groups, but these small groups were

TABLE II.—DISTRIBUTION OF EXTRAVERSION-INTROVERSION SCORES BY SEX, SCHOOL GRADE, AND RELIGIOUS SECT

Sex	School Grade	Sect	Mean Score	SE Sample	SE Mean	Diff. in Means	SE Diff.	D/SE Diff.
Boys	V–VI	Amish	19.19	4.41	0.99	2.52	1.39	1.81
		Non-Amish	21.71	4.36	0.98			
	VII–VIII	Amish	18.76	3.36	0.75	1.76	1.02	1.73
		Non-Amish	20.52	3.51	0.69			
Girls	V–VI	Amish	19.26	3.76	0.80	2.19	1.25	1.75
		Non-Amish	21.45	4.19	0.96			
	VII–VIII	Amish	18.40	3.36	0.89	3.38	1.12	3.02
		Non-Amish	21.78	2.80	0.68			
Both	V–VIII	Amish	18.95	3.77	0.42	2.34	0.59	3.97
		Non-Amish	21.29	3.81	0.41			

used because the authors of the inventory give norms by sexes for grades IV, V, and VI and for grades VII, VIII, and IX. Thus comparison of the groups in the present study could be made with the large groups on which the norms were based. Furthermore, it was thought that in a comparison of fifth- and sixth-grade Amish and non-Amish children with those of seventh and eighth grade, such children might show increasing differences with age. In addition, comparisons were made between the scores of all Amish children and those of all non-Amish children.

In Table I it will be noted that for every group the mean scores on Ascendance-Submission for Amish children are lower than for non-Amish children and that the difference between the mean scores of all Amish and all non-Amish children is statistically significant. On this inventory low scores indicate a "submissive, retiring type of child."

It will be noted that for every group in Table II mean scores on Extraversion-Introversion for Amish children are lower than the mean scores for non-Amish children and that the difference between mean scores for all Amish and all non-Amish children is statistically significant. On this inventory low scores are indicative of introversion.

TABLE III.—DISTRIBUTION OF EMOTIONAL STABILITY SCORES BY SEX, SCHOOL GRADE, AND RELIGIOUS SECT

Sex	School Grade	Sect	Mean Score	SE Sample	SE Mean	Diff. in Means	SE Diff.	D/SE Diff.
Boys	V–VI	Amish	25.10	6.75	1.51	0.33	2.10	0.16
		Non-Amish	25.43	6.52	1.46			
	VII–VIII	Amish	27.67	2.93	0.66	0.22	1.07	0.21
		Non-Amish	27.89	4.27	0.84			
Girls	V–VI	Amish	24.00	5.98	1.28	0.55	1.92	0.28
		Non-Amish	23.45	6.24	1.43			
	VII–VIII	Amish	25.40	4.38	1.17	0.07	1.69	0.04
		Non-Amish	25.33	4.98	1.21			
Both	V–VIII	Amish	25.51	5.48	0.62	0.21	0.88	2.39
		Non-Amish	25.72	5.75	0.63			

In Table III it is of interest to note that on Emotional Stability differences between mean scores for Amish and non-Amish children are very slight, in the cases of boys the Amish having slightly lower mean scores than the non-Amish, but in the cases of girls the Amish having slightly higher mean scores than the non-

Amish. Low scores on this inventory are indicative of emotional instability. The fact that there are such very slight differences for Amish and non-Amish children in mean scores on Emotional Stability may possibly suggest the additional importance of such differences as were found for Ascendance-Submission and for Extraversion-Introversion.

The principal interest in this study has been in a comparison of Amish and non-Amish children attending the same schools, but as a point of further interest both groups have been compared with the norm groups as indicated in the manual accompanying the inventory. The authors of the inventory have published tables of percentile ranks for scores on each trait by sex and grade groups. By a simple process of direct interpolation in these tables approximate percentile ranks have been estimated for various groups in the present study as indicated in Table IV. In no aspect of personality as measured by the inventory are Amish children as high as the 50th percentile. In Emotional Stability all groups of both Amish and non-Amish children are below the 50th percentile, and it is of interest to note that in all groups children in the seventh and eighth grades show greater instability than children in the fifth and sixth grades. Except for non-Amish girls in the fifth and sixth grades, all groups are below the 50th percentile in Ascendance-Submission. Except for boys in the seventh and eighth grades all groups of non-Amish children are above the 50th percentile in Extraversion-Introversion whereas, as has been noted, all Amish groups are below this percentile.

Inasmuch as differences in Ascendance-Submission and Extraversion-Introversion were shown by Amish and non-Amish groups, a further analysis was made of the responses given to the thirty-five statements in each of these sections. The responses made by all Amish boys were compared with the responses made by all non-Amish boys and the same was done for girls. In each case the percentage of each group responding to each statement by "same" was computed. Only the three statements showing greatest percentage differences between Amish and non-Amish subjects are reported here.

In the section of the inventory measuring Ascendance-Submission, the statement, "I don't mind when other children get ahead of me in line," was answered "same" by thirty-two

(76.2 per cent) Amish boys, but this answer was given by only twenty-four (50.0 per cent) non-Amish boys. The statement showing the next greatest difference was, "I talk back to a friend who is 'bossy.'" Fourteen (33.3 per cent) Amish boys answered "same" to this statement, but twenty-eight (58.3 per cent) non-Amish boys agreed. "I stick to what I've said even if other children don't like it," was marked "same" by eighteen (42.9 per cent) Amish boys and by thirty-one (64.6 per cent) non-Amish boys. The significance ratios for the differences in percentages for these three statements are 2.05, 1.59, and 1.50, respectively.

In the same section of the inventory ten (26.3 per cent) Amish girls agreed with the statement, "I talk back to a friend who is 'bossy,'" and twenty-four (63.2 per cent) non-Amish girls agreed. The statement, "I am among the first to yell at a game," was agreed to by eight (21.1 per cent) and twenty (52.6 per cent) Amish and non-Amish girls, respectively. "If there are pieces of salt in my ice cream, I tell the storekeeper about it," was answered "same" by sixteen (42.1 per cent) Amish girls but by only four (10.5 per cent) non-Amish girls. (Amish people have a reputation for driving hard bargains.) The significance ratios for the differences in percentages for these three statements are 1.87, 1.90, and 1.60, respectively.

Of the statements in the Extraversion-Introversion section of the inventory the one showing the greatest differentiation between Amish and non-Amish boys was, "I do not make friends easily." Twelve (28.6 per cent) Amish boys agreed with this statement whereas only two (4.2 per cent) non-Amish boys agreed. "I keep quiet when I am with other people," was agreed to by twenty-three (54.8 per cent) Amish and seventeen (35.4 per cent) non-Amish boys. Twenty-eight (66.7 per cent) Amish boys answered "same" to the statement, "I like to read poetry," but only twenty-three (47.9 per cent) non-Amish boys agreed with this statement. The significance ratios for the differences in percentages for these three statements are 1.27, 1.24, and 1.37, respectively.

For girls the statement which differentiated most between Amish and non-Amish subjects in the Extraversion-Introversion section was, "I like to spend money"; fifteen (39.5 per cent) Amish and thirty (78.9 per cent) non-Amish girls agreeing with

the statement. (The Amish are reputed to be very frugal.) "I feel at home at parties," was agreed to by nine (23.7 per cent) Amish and twenty-four (63.2 per cent) non-Amish girls. "I like to belong to clubs," was agreed to by twelve (31.6 per cent) and twenty-six (68.4 per cent) Amish and non-Amish girls, respectively. The significance ratios for the differences in percentages for these three statements are 2.70, 2.28 and 2.27, respectively.

TABLE IV.—PERCENTILE RANKS BY SEX AND SCHOOL GRADE OF AMISH AND NON-AMISH CHILDREN ON THE BASIS OF NORMS PUBLISHED FOR PINTNER'S *Aspects of Personality*

Sex	School Grade	Sect	Ascendance-Submission	Extraversion-Introversion	Emotional Stability
Boys	V–VI	Amish	29	37	44
		Non-Amish	40	65	46
	VII–VIII	Amish	6	24	33
		Non-Amish	22	37	35
Girls	V–VI	Amish	33	38	37
		Non-Amish	64	63	34
	VII–VIII	Amish	16	15	25
		Non-Amish	25	51	24

SUMMARY

In the present study it has been found that belonging to a religious sect which insists upon unusual customs, including the wearing of a garb, does seem to produce personality patterns which are somewhat different from the personality patterns of other children, at least in so far as such patterns are measured

by a personality inventory. Some tendency has been found for Amish children to be more submissive and introverted than non-Amish children, as is shown by their willingness to let other children get ahead of them in a line, to let other children 'boss' them without talking back, by difficulty in making friends easily, by not liking to belong to clubs, by not feeling at home at parties, and so on. This tendency toward submissiveness and introversion was found not only on the basis of comparison with other children attending the same schools but on the basis of norms as published by the authors of the personality inventory. On the other hand, Amish children seem to be as well adjusted emotionally as non-Amish children attending the same schools although both groups have been found to be more unstable than the norm groups for the inventory. There was some indication that for both Amish and non-Amish groups emotional instability increased with age.

Further research on this problem is needed. In the present study approximately half of the children in each school were Amish and half were non-Amish. It would be interesting to compare school groups in which only a very small percentage of the children were Amish with school groups in which a very large percentage of the children were Amish. In the present study the schools selected were in rural districts, and there was not as contrasting a difference in dress and customs between Amish and non-Amish children as there would be in most urban schools. Such a comparison would be interesting but difficult to make because the Amish are predominately an agricultural people. Possibly differences in attitudes would be greater in adolescence than has been shown in the present study of elementary-school children, but such a study would be difficult to make, for, as a rule, the Amish do not permit their children to attend high school. It would probably be difficult to secure the coöperation of Amish young people who were not in school. Measurement of personality other than by means of an inventory would be valuable. Obviously larger groups of subjects are needed.

HOW IS INTELLECTUAL PERFORMANCE RELATED TO SOCIAL AND ECONOMIC BACKGROUND

RICHARDS C. OSBORN

Kalamazoo College

INTRODUCTION

The question, "Is a man smart because he picked a good environment or good parents?" has provoked much discussion among educators, sociologists, psychologists, and biologists. Investigations have taken three general forms: first, the tracing of superiority or inferiority in families; second, the studies of the constancy of the IQ and the counterclaims, especially in the Iowa Studies, of a flexible IQ, influenced markedly by environmental conditions; and, third, the measurement of the socio-economic status of different groups in relation to their intelligence. The present investigation falls into this last group of studies of social and economic background.

Numerous studies have related differences in socio-economic status to differing levels of intelligence and achievement. Most of these have used either occupational classification or a weighted scale of factors considered of chief significance in measuring status.[1]

The majority of general investigations has shown that the background factors for persons at the top of the IQ scale are superior to the ones exhibited by those at the bottom of that scale. One of the investigations, however, has shown a surprisingly small superiority for those in the better environment. A comparison was made over four years between relief and non-relief children in the schools of Wichita, Kansas. It was discovered that 30 per cent of the relief and 34.5 per cent of the non-relief children were above average in grades, while 41.3 per cent of the relief and 53.65 per cent of the others were above average in intelligence.

---

[1] Sims, Verner Martin: *The Measurement of Socio-economic Status.* Public School Publishing Company, Bloomington, Illinois, 1928.

Chapin, F. Stuart: *The Measurement of Social Status.* The University of Minnesota Press, Minneapolis, 1933.

Neff, Walter S.: "Socio-economic Status and Intelligence: A Critical Survey." *Psychological Bulletin,* Vol. xxxv, 1938, pp. 727–754.

In other words, the relief children were inferior on the average, but there was a tremendous amount of overlapping.[1]

Several surveys have been made of the influence of status on the test performance of college students. They have been somewhat brief but the findings can be summarized thus: "There is evidently a tendency for freshmen in the higher socio-economic centiles to have more intelligence (as determined by intelligence tests) than freshmen in the lower centiles have."[2] Although there was a definite relation between intelligence and background, a very low correlation was found to exist between the two.

The National Resources Committee has presented a comparison, for various states, of certain background factors such as newspaper circulation, education, and income, with Army intelligence ratings. The rather striking conclusion drawn is that: "Intelligence-test scores for individuals brought up in similar environments as regards home, community, and school influences, are strongly influenced by hereditary factors. The same tests, however, when applied to groups matured in very different environmental conditions, exposed to diverse family traditions, and trained in different schools, may be used to measure differences in cultural opportunity."[3]

This statement of the institutional approach to the problem, suggests that differences in intelligence within homogeneous groups may possibly be hereditary, but that average differences among diverse groups are presumed to be principally environmental.

STATEMENT OF THE PRESENT STUDY

This paper attempts to show the relation between performance and cultural backgrounds. To do this a questionnaire on social and economic backgrounds was given to freshmen in three different colleges in the Middlewest. The data covered are summarized in Table I.

---

[1] Croft, Albert E.: "The Ability of Relief Children." *American Sociological Review*, Vol. v, 1940, pp. 185–192.

[2] Cuff, Noel B.: "Relationship of Socio-economic Status to Intelligence and Achievement." *Peabody Journal of Education*, Vol. xi, No. 3, p. 109. The author used the American Council Psychological Test to measure intelligence.

[3] National Resources Committee: *The Problems of a Changing Population*. United States Government Printing Office. Washington, 1938, p. 232.

TABLE I.[1]—AREAS OF BACKGROUND COVERED BY THE QUESTIONNAIRE

DATA CONCERNING THE STUDENT	DATA CONCERNING THE FAMILY BACKGROUND
1. Age	1. Age of parents
2. Sex	2. Country of parent's birth
3. Birthplace	3. Schooling of parents
4. Residence	4. Employment of parents
5. Extent of travel	5. Parental income
6. High-school background	6. Type of house
7. Means of college support	7. Modern conveniences
8. Vocation planned	8. Reading materials
	9. Church attendance
	10. Mobility of family

[1] Tables V and VI give the summaries of the results of the questionnaire.

The factors of background presented in the questionnaire are not all-inclusive. They are those objective features about which the students were most likely to be conversant. No attempt is made to reveal attitudes, personality ratings, or like subjective measurements. Such estimates are open to much more varied interpretation than the ones chosen. The factors selected are, for the most part, those that have been connected, either in popular opinion or scientific research, to performance or to intelligence—usually being considered of causal significance.[1] No attempt is made here to prove causal relationship; an effort is made merely to point out the differences in background which exist among students with differing qualities of performance.

Briefly, these are the groups to which the questionnaire was given: College *A* is a small denominational liberal arts school and the whole freshman class participated. College *B* is a teachers college. Here the questionnaire was sent to a representative sample of the freshman class, selected by the educa-

[1] Two widely publicized books on happiness in marriage have shown that differences in background factors exist as between happily and unhappily married couples, and the authors considered these background differences as of causal significance. No such attempt to explain performance on the basis of environmental causation is deemed advisable. There is an effort made merely to show how the social and economic background of the superior students is similar to and how it differs from the background of the inferior ones.

tional research clinic of the school. College *C* is a teachers college. At this school also the questionnaire was given to the entire freshman class, which included four-year, two-year intermediate, two-year elementary, one-year rural, and twelve-week normal training groups.[1]

The Psychological Examination of the American Council on Education had previously been given to the students of Colleges *A* and *C*.[2] To those of College *B* the Ohio College Association Examination had been presented.

The students[3] of each school were divided into fifths, ranging from high to low, based on their performance in these tests. An attempt was then made to see what differences in background factors existed for each level of performance. Table II shows the distribution of students approximately into fifths for each school as a whole. These are also subdivided further to subdivide College *C* and to separate town from farm students.[4]

Table III shows the high and low scores, the quartiles, and the means for the different college groups. It is presented in terms of percentile rank based on the national norms for the tests involved.[5]

It becomes apparent from Table III that there are considerable differences in average intellectual performance among the groups and that wide differences exist within each group.

This study will attempt to answer the following questions with respect to the groups included in the questionnaire:

(1) What are the differences on the average in background factors of the groups?[6] For example, how do the average age of

---

[1] On all questionnaires there is a problem concerning the honesty of the students' answers. It is the opinion of the writer that the students, in all but a few obvious instances, put down what they thought was correct. Any errors in their statements arise through ignorance rather than malicious intent.

[2] Correlations of .68–.72 between the psychological examination and the first-semester grades have been obtained consistantly for College *C* freshmen.

[3] This refers to the freshmen in each case mentioned.

[4] The separation of farm from town students was necessary because of the difference in background characteristics of each.

[5] No farm students for College *A*.

[6] The different groups are College *A*; College *B* town; College *Ca*, *Cb*, *Cc*, *Cd*, and *Ce* town; College *B* farm; College *Ca*, *Cb*, *Cc*, *Cd*, and *Ce* farm.

students, years of schooling of fathers, etc., compare for Colleges *A*, *B*, *Ca*, *b*, *c*, *d*, and *e*, in both town and farm classifications?

TABLE II.—DISTRIBUTION OF STUDENTS BY NUMBERS IN EACH GROUP[1]

	Total Numbers	High Fifth	Second Fifth	Middle Fifth	Fourth Fifth	Low Fifth
All Students						
		1	2	3	4	5
College *A*—Four Year	136	27	27	28	27	27
College *B*—Four Year	100	22	21	21	14	22
College C—Four Year	275	75	61	57	40	42
College *C*—All Groups	650	130	130	130	130	130
Town Students						
College *A*—Four Year	136	27	27	28	27	27
College *B*—Four Year	73	15	17	15	10	16
College *C a* Four Year	208	62	44	48	25	29
b Two Year Intermediate	36	7	7	6	11	5
c Two Year Elementary	94	18	23	10	20	23
d One Year Rural	15	..	6	2	4	3
e Twelve Week Normal	17	..	1	1	7	8
Farm Students[2]						
College *B*—Four Year	27	7	4	6	4	6
College *C a* Four Year	67	13	17	9	15	13
b Two Year Intermediate	55	12	8	18	10	7
c Two Year Elementary	63	7	9	13	17	17
d One Year Rural	41	6	4	15	8	8
e Twelve Week Normal	54	5	11	8	13	17

[1] The superior fifth of the students is found in column 1 and the poorest fifth in column 5 in all of the tables. It will be noted that the subdivisions of each college do not necessarily contain equal numbers, since some groups have a large concentration of good students and *vice versa*.

[2] College *A* percentiles are for the 1940 hand-scored edition of the American Council Psychological Test. College *C* percentiles are from the 1938 machine-scored edition of the same test. College *B* percentiles are desired from the 1940 edition of the Ohio College Association Test.

(2) What is the variance in background within each freshman group—for example, the differences existing in average age, schooling of fathers, etc., between the top two-fifths and the

TABLE III.—PERCENTILE RANK OF STUDENTS
(Based on National Norms)

	High	Q3	Mean	Q1	Low
ALL STUDENTS					
College *A*—Four Year.....	99.5	80	62	36	10
College *B*—Four Year.....	98	71	50	30	1
College *C*—Four Year.....	99.5	73	52	30.5	1
—All Groups.....	99.5	64	44	24	1
TOWN STUDENTS					
College *A*—Four Year.....	99.5	80	62	36	10
College *B*—Four Year[1].....					
College *C a* Four Year.....	99.5	74	54	33.5	3
b Two Year Intermediate.....	96.5	61.5	46	24	9
c Two Year Elementary.....	95.5	61.5	43	18.5	3
d One Year Rural.	70	63	41	20	8
e Twelve Week Normal.....	67	29	18	5	3
FARM STUDENTS					
College *B*—Four Year[1].....	95.5	67	43	23	1
College *C a* Four Year.....	91.5	68.5	47	26.5	3.5
b Two Year Intermediate.....	93	55.5	37	17	2
c Two Year Elementary.....	87	51	40	23	4
d One Year Rural.	87	51	40	23	4
e Twelve Week Normal.....	86	53	32	17	1.5

[1] Because of the way in which the questionnaire was given to College *B* there was no way to compare the percentile rank of town and farm students.

bottom two-fifths within Colleges *A*, *B*, *C*, and subdivisions of the latter?

(3) What major conclusions can be drawn and what generalizations made on the basis of the data collected?

A study of Table IV showing the years of schooling of fathers will illustrate further. A comparison of the averages vertically shows the differences in schooling of fathers among the groups. The comparison of the figures horizontally will show the schooling of fathers in the different fifths within each group.

TABLE IV.—SCHOOLING OF FATHERS (YEARS)

	Averages	High Fifth	Second Fifth	Middle Fifth	Fourth Fifth	Low Fifth
All Students						
College *A*—Four Year	12.5	13.7	13.4	11.6	11.1	12.7
College *B*—Four Year	10.6	10.7	10.1	10.6	11.0	10.8
College *C*—Four Year	11.2	12.3	10.9	11.1	10.7	10.8
All Groups	10.5	11.3	10.9	9.7	10.6	10.1
Town Students						
College *A*—Four Year	12.5	13.7	13.4	11.6	11.1	12.7
College *B*—Four Year	11.0	10.9	10.9	11.5	11.0	11.5
College *C a* Four Year	11.5	12.2	11.6	10.4	12.3	11.2
b 2 Yr. Inter	10.6	11.1	10.3	10.5	10.0	11.2
c 2 Yr. Elem	11.7	13.0	11.9	10.9	11.4	11.6
d 1 Yr. Rural	9.1	....	9.0	7.5	9.5	10.0
e 12 Wk. Norm	8.6	....	10.0	7.0	8.3	8.8
Farm Students						
College *B*—Four Year	9.6	10.6	7.0	8.3	10.7	10.7
College *C a* Four Year	10.0	10.8	10.9	8.9	10.0	8.6
b 2 Yr. Inter	9.7	9.1	7.1	9.4	10.9	11.4
c 2 Yr. Elem	10.1	9.6	9.8	9.5	11.5	9.5
d 1 Yr. Rural	8.9	7.2	10.5	9.1	9.9	8.3
e 12 Wk. Norm	8.4	8.2	9.1	8.5	7.5	8.4

DIFFERENCES IN BACKGROUND FACTORS AMONG GROUPS

The differences in cultural background among groups is quite marked. In most cases College *A* exhibits superiority in those factors of background considered to be advantageous, with the subdivisions of College *C* showing up the most poorly. By referring again to Table III it becomes evident that there are decided differences in intellectual performance, especially in towns, as among the different classifications of students. Although differences in average test performance are greater in the different fifths within each group than among groups, the differences in background factors as among groups are greater than the differences in background factors within each group. The background within any one school, or sub-classification of a school, was more homogeneous than that among schools.[1]

[1] An exception to this range of background and performance is found in the two year intermediate (*Cb*) group of students. In the case of both town and farm members they show up fairly well in performance, although having a set of background factors inferior to those of the two-year elementary (*Cc*) students in a large number of items.

The data on the differences in background factors among groups is shown in Table V. This table indicates that those groups which were superior in test performance had a significantly different background as shown by the higher education of parents; they live in larger towns, come from smaller families, have parents with larger incomes, own more and newer modern

TABLE V.—DIFFERENCES IN BACKGROUND FACTORS AMONG GROUPS[1]

	Colleges and Groups						
	Town Students				Farm Students		
	A	*B*	*Ca*	*Cc*	*B*	*Ca*	*Cc*
Average Age of Students (Years)	17.9	18.8	18.3	18.1	18.4	18.3	17.8
Per Cent Male	47.1	50.7	62.7	0.0	33.3	17.9	0.0
Born in Same State as School (Per Cent)	64.0	82.2	83.2	89.5	77.8	95.5	95.1
Size of Home Town (Median)	59,000	8,500	6,300	1,300	2,000	925	1,100
Average Age of Father (Years)	48.9	49.6	49.2	51.4	51.4	50.2	50.1
Average Age of Mother (Years)	45.9	46.4	46.8	51.4	47.9	47.5	46.1
Fathers Born in U. S. (Per Cent)	87.8	91.5	85.2	95.6	77.8	90.8	95.1
Mothers Born in U. S. (Per Cent)	94.6	91.4	94.6	95.5	77.8	92.1	95.2
Schooling of Fathers (Years)	12.5	11.0	11.5	11.7	9.6	10.0	10.1
Schooling of Mothers (Years)	12.6	11.1	11.5	11.9	10.7	10.4	10.5
Children in Family (Number)	2.6	3.3	3.3	3.1	4.3	4.3	4.1
Fathers Living (Per Cent)	97.0	91.8	89.6	89.4	100.0	94.0	96.8
Mothers Living (Per Cent)	96.9	94.4	94.7	95.7	96.3	87.7	95.2
Depression Unemployment (Per Cent)	19.8	23.9	17.3	10.0	20.0	1.6	3.4
Average Yearly Income ($)	3,431	2,174	2,115	1,549	1,353	1,973	1,716
Median Yearly Income ($)	2,900	1,950	1,550	1,300	1,150	1,800	1,800
Mothers Employed (Per Cent)	12.0	16.4	14.2	14.1	0.0	1.8	1.7
Year Automobile Purchased	38.3	36.8	36.3	35.7	36.0	36.5	35.5
Year Radio Purchased	36.9	36.7	35.8	35.1	36.2	36.3	36.8
Homes Having Telephone (Per Cent)	90.3	80.3	89.4	89.1	50.0	87.3	88.9

TABLE V.—*Continued*

	Colleges and Groups						
	Town Students				Farm Students		
	A	*B*	*Ca*	*Cc*	*B*	*Ca*	*Cc*
Homes Having Running Water (Per Cent).....	99.2	98.6	94.5	90.3	50.0	41.9	48.4
Homes Having Electricity (Per Cent).....	100.0	100.0	99.0	100.0	81.5	68.9	69.8
Homes Having Kitchen Sink (Per Cent).....	99.3	100.0	99.0	98.9	81.5	87.1	85.5
Mechanical Refrigerator (Per Cent).....	89.4	78.1	59.5	58.2	53.8	31.6	45.8
No Refrigeration in Home (Per Cent).....	0.8	5.5	5.5	13.2	11.5	38.6	37.3
Daily Newspaper in Home (Per Cent).....	99.2	93.0	97.1	93.6	85.2	95.1	98.4
Family Magazines (Number).....	5.8	4.7	4.9	4.9	5.7	6.8	5.9
Books in Family Library (Number).....	184	139	108	113	96	94	85
Parents Use Public Library (Per Cent).....	75.0	79.7	75.5	72.4	53.8	45.8	48.3
Church Attendance (Times a year).....	39.1	34.6	42.6	48.3	29.8	46.7	49.7
Parents' Church Attendance (Year).....	37.5	35.1	42.3	46.1	28.4	47.2	45.9
Times Movies Attended (Per mo.).....	3.5	3.6	3.4	3.5	2.6	3.1	2.9
Students Dance (Per Cent)	92.4	89.0	82.0	79.6	85.2	52.5	66.1
Students Play Cards (Per Cent).....	84.0	84.9	82.4	76.1	81.5	61.5	77.4
Time on Daily Paper (Minutes).....	21.8	18.4	23.6	20.1	18.7	26.4	19.3
Students Lived in Only One State (Per Cent)[1]..	70.4	73.2	72.4	81.5	77.8	92.2	87.3
Size of Graduating Class..	245	107	74	35	55	24	28
Ranked in Upper Third (Per Cent).....	88.2	74.6	62.3	70.3	68.2	74.6	70.4
Employed during High School (Per Cent).....	36.1	60.3	52.7	28.6	40.7	50.0	19.4
Directly to College from H. S. (Per Cent).....	83.3	77.5	71.3	76.6	85.2	71.2	79.4
Parents Wholly Supporting (Per Cent).....	54.1	31.5	36.4	57.0	29.6	33.3	51.6

[1] In order to conserve space the writer has included only part of the subdivisions of College *C*.

conveniences, enjoy more literary life, and travel more extensively. In short, they had more of those features which are usually considered as constituting superior social and economic background.

COMPARISON OF BACKGROUND FACTORS WITHIN EACH GROUP

This study shows a tendency for the poorer students within each group (those in the lower two fifths) to have some inferior background characteristics. These are both less numerous and less extensive, however, than were the variations in background as among groups. Of the list of factors appearing in Table V, showing differences among groups, only a few show important variation as between the better and poorer students within each group.

The students in the lower fifths had parents with the same average education,[1] came from the same size of town, had the same size of family, and enjoyed other similar background char-

TABLE VI.—DIFFERENCES WITHIN GROUPS
(Four-year Town Students)

	Colleges			Colleges			Colleges		
	A	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*
	Father Engaged in a Profession (Per Cent)			Parents Use Public Library (Per Cent)			Students Employed while in High School (Per Cent)		
High Fifth	29.2	11.1	21.1	78.3	66.7	80.0	32.0	60.0	40.3
Second Fifth	29.2	17.7	14.3	77.8	76.5	88.1	36.0	41.2	41.9
Middle Fifth	8.4	14.3	10.6	80.0	85.7	68.9	46.4	80.0	65.3
Fourth Fifth	12.5	0.0	14.7	73.1	90.0	77.3	26.3	70.0	58.3
Low Fifth	19.2	20.0	10.3	60.9	81.3	55.6	36.0	56.3	75.0
	Median Income ($)			Attend Movies (Times per Month)			Students for Whom No Time Elapsed between High School and College (Per Cent)		
High Fifth	2950	1900	1850	3.4	3.8	3.1	80.0	85.7	75.8
Second Fifth	2450	2100	1650	3.0	2.7	3.5	80.0	75.0	82.9
Middle Fifth	3500	1950	1550	3.4	3.4	3.5	92.9	73.3	64.6
Fourth Fifth	2500	2200	1200	3.9	4.5	3.6	76.0	70.0	70.8
Low Fifth	2500	1600	1600	3.9	3.8	3.3	86.4	81.3	55.5

[1] Except for College *A*, in which the fathers at the top had more education.

TABLE VI.—*Continued*

	Colleges A	Colleges B	Colleges C	Colleges A	Colleges B	Colleges C	Colleges A	Colleges B	Colleges C
	Mothers Employed (Per Cent)			Minutes Devoted to Daily Newspaper			Wholly Supported by Parents (Per Cent)		
High Fifth	20.8	20.0	18.6	20.5	17.0	24.7	51.9	33.3	36.1
Second Fifth	8.0	11.8	17.9	19.0	19.7	24.9	60.0	35.3	46.3
Middle Fifth	18.5	20.0	10.9	20.0	16.3	21.2	39.3	36.3	32.7
Fourth Fifth	4.1	30.0	9.5	23.6	21.8	30.0	55.6	22.2	33.3
Low Fifth	8.0	16.3	8.0	24.0	18.4	18.2	63.0	31.3	35.7
	Magazines (per Family)			Times Parents Have Moved			Vocation Planned in Profession or Research (Per Cent)		
High Fifth	5.8	5.0	5.4	2.7	3.5	3.8	87.5	25.0	26.0
Second Fifth	5.2	5.5	5.3	3.7	5.5	4.2	50.0	25.0	20.0
Middle Fifth	5.8	3.9	4.1	3.7	3.6	3.7	29.4	13.3	14.0
Fourth Fifth	5.4	4.1	5.0	2.0	4.4	3.9	31.2	20.0	10.0
Low Fifth	7.0	4.8	4.4	3.0	2.9	3.5	43.8	17.6	4.0
	Number of Books in Family Library			Ranked in Upper Third of High School Class (Per Cent)			Vocation Planned in Teaching (Per Cent)		
High Fifth	153	146	116	90.0	100.0	80.0	0.0	75.0	26.0
Second Fifth	200	174	128	86.7	83.3	63.3	0.0	50.0	54.0
Middle Fifth	218	99	98	95.0	75.0	66.7	17.6	66.7	64.0
Fourth Fifth	168	84	103	90.0	66.7	42.9	6.3	60.0	74.0
Low Fifth	174	120	76	76.9	46.2	33.3	12.5	58.8	70.0

acteristics. Only small average differences appear in those factors that we call modern conveniences, such as radios, refrigerators, electricity, automobiles, etc.

The factors showing the greatest differences in background as among fifths are illustrated in Table VI.[1] From a study of this table it becomes apparent that the parents of those in the superior fifths had a higher median income, better reading facilities, and contributed more financial support. The superior students had less delay between high-school and college enrollment. They came in larger numbers from the professional class and more of them planned to remain in that class. For the other items in

[1] In place of showing data for all the groups, the four-year town students from each school were selected for Table VI.

the study the superior groups appeared slightly better in some and made a poorer showing in others.[1]

CONCLUSIONS

In drawing inferences from statistical data there is always danger of universalizing particular conclusions, which may apply only to the data at hand. The writer does not wish to suggest that the groups studied here are typical of all classifications. College students represent a select category. The proportion of those from the professional and business classes is much larger than is the case for the population as a whole. The students coming from the unskilled working class are often the brighter ones from the better homes in that class. In fact, this selective process is apparent even in the high school.

It is nevertheless possible to state certain conclusions.

(1) From this study it becomes clear that the cultural differences within a group (or school) background are much less than the differences found as among the various groups. Within any one group of students the background of the poorest is not much different from that of the best. Comparing groups or schools, however, we find wide differences in background.

(2) We seem to have layers of fairly homogeneous background for each group. Within these layers the variation in performance is rather wide—varying in most cases from near the top percentiles of national distribution to the bottom ones. The different layers are represented by different average levels of performance, which are accompanied by different average background factors. Within groups the performance found in the different fifths has wide variation; the background, but little. Comparing groups, especially on the farms, the differences in performance are small but the background differences are quite marked.

If we ignore these levels of background the conclusions will be confused. If, for example, we do not segregate the different groups, but lump them all into equal fifths based on test performance, the top fifth would be composed primarily of four-year students, and College *A* would dominate. The last fifth would be made up of small town and farm one-year rural and twelve-

[1] In summarizing the intra-school differences it is possible to say that, with respect to the students investigated, no differences of consistent significance existed in the majority of the factors studied.

week normal girls. This would not give a clear and concise picture of the actual situation.

(3) It appears that within groups a superior background has not produced superior performance; nor the superior performance, a better background. If the different backgrounds within each group did not produce superior students within that group, can we explain the differences among groups on this basis? If we were trying to improve the performance of College *B* and College *C* students, would we merely move them to the environment of College *A*? In any case, we already have some students in College *A* who, in spite of their superior background, are inferior to many in Colleges *B* and *C*.

(4) So far as this study is concerned, the results are not such as to permit the acceptance of environmental determinism as the chief explanation of the differences in performance within groups. Among groups the environment has no doubt had some effect—but among farm groups the differences in performance are small, in spite of considerable background variation. Apparently those who ignore or under-estimate the importance of hereditary differences in intellectual performance still have a burden of proof.

(5) To indicate that differences in environmental factors have not appeared sufficient to act as major determinants of differences in intellectual performance is not to deny the significance of nurture in the development of the whole personality of a person. We do not live in an isolated area of intellectual performance. Our whole societal structure is a powerful institutional determiner of our ideas, attitudes, and conduct. Some areas and some groups produce crime and delinquency, others produce high moral codes and superior ethical conduct. It is desirable, however, to learn as much as possible about the changes which environment can foster and to direct our principal efforts into those channels.

SELECTED BIBLIOGRAPHY

Glenn Myers Blair: *Mentally Superior and Inferior Children in the Junior and Senior High School.* Teachers College, Columbia University, Contributions to Education No. 766, New York, 1938.

Ruth Byrns, and V. A. C. Henmon: "Parental Occupation and Mental Ability." *Journal of Educational Psychology*, Vol. XXVII, April, 1936, pp. 284–291.

F. Stuart Chapin: *The Measurement of Social Status.* The University of Minnesota Press, Minneapolis, 1933.

George Sylvester Counts: *The Selective Character of American Secondary Education.* Supplementary Educational Monograph No. 19, University of Chicago Press, University of Chicago, 1922.

Albert E. Croft: "The Ability of Relief Children." *American Sociological Review*, Vol. v, 1942, pp. 185–192.

Noel B. Cuff: "Relationship of Socio-economic Status to Intelligence and Achievement." *Peabody Journal of Education*, Vol. xi, No. 3. P. 109.

Minerva F. Desing: *The Relation of Pupil Achievement Gain to Certain Personal and Environmental Elements.* Doctoral Dissertation, University of Pennsylvania, 1940.

L. C. Glass: "The Relation of the 'Intelligence' of College Students to the Occupation of the Parents." *Eugenical News*, Vol. xxi, January, 1936, pp. 1–2.

Charles E. Halley: "The Relationship Between Persistence in School and Home Conditions." National Society for the Study of Education. *Fifteenth Yearbook*, Part II, 1916.

E. J. Howell: "Economic and Social Background of Freshman Engineers and Academic Achievement." *Bulletin of The American Association of Collegiate Registrars*, Vol. x, 1935, pp. 262–270.

M'Ledge Moffett: *Social Background and Activities of Teachers College Students.* Teachers College, Columbia University, Contributions to Education, No. 375, 1929.

National Resources Committee: *The Problems of a Changing Population.* United States Government Printing Office, Washington, 1938.

Walter S. Neff: "Socioeconomic Status and Intelligence: A Critical Survey." *Psychological Bulletin*, Vol. xxxv, 1938, pp. 727–754.

Lillian Portenier: "Aspects of Socio-economic Status Which Differentiate University Students." *Journal of Applied Psychology*, Vol. xxii, June, 1938, pp. 261–270.

Ora Edgar Reynolds: *Social and Economic Status of College Students.* Teachers College, Columbia University, Contributions to Education, No. 272, 1927.

Verner Martin Sims: *The Measurement of Socio-economic Status.* Public School Publishing Company, Bloomington, Illinois, 1928.

Ruth Strang: *Behavior and Background of Students in College Secondary School*, Harper and Brothers, New York, 1937.

FACTORS RELATED TO HIGH-SCHOOL ACHIEVEMENT

VIOLA AMES

Purdue University

For many years educators have been interested in the prediction of high-school achievement from intelligence scores. The present study is an attempt to isolate certain other personality traits which may be related to achievement. That intelligence alone does not account for all variation in scholastic achievement is evidenced by many studies, a few of which are reported here.

Byrons and Henmon,[1] in a study of long range prediction of college success, correlated the scores on the National Intelligence Test, given to Grades IV to VIII in 1920, with the average achievement of the same pupils at various school levels. The correlations between average grade achievement and intelligence as found at various levels of school progress are given below:

At the tenth year of school	$r = .397$
At the fourth year of high school	$r = .426$
At the first year of college	$r = .454$
At the fourth year of college	$r = .370$

Kohn[2] reports that Pintner, upon compiling the correlations of intelligence and achievement as reported by fourteen authors, found a range of correlations from $r = .28$ to $r = .60$ with $r = .46$ as a mean.

In his study of the relationship between achievement and intelligence, as measured during the elementary school, Adams[3] found the median IQ's at the various school levels to be:

College level	112.16
Bachelor's degree level	115.0
Master's degree level	119.5

He also reports four males and nine females with IQ's ranging from 84 to 99 who received bachelor's degrees.

Keys[4] quotes Portenier as having found in a group of ninth-grade students, classified as dull, a range of 67 to 108 in I.Q. Of these:

68 per cent completed high school
50 per cent completed one year of college
32 per cent completed four years of college

The following figures are from Byrons and Henmon's[1] study:

Median IQ of high-school seniors	107.7
Mean IQ of college freshmen	115.5
Mean IQ of college graduates	118.3

Two of the college graduates in this study had IQ's below 95.

The above studies, along with many more, indicate that the range of IQ's at any school level is large, and the correlation of intelligence and achievement in school far from perfect.

Using five intelligence tests (Otis Intelligence, Terman Group Intelligence, Henmon-Nelson Intelligence, National Intelligence, and Kuhlmann-Anderson Intelligence) and the Ohio State University Psychological Examination, Hartsom and Sprow[6] found no significant difference in the value of these tests for predicting college scholastic achievement. The correlation of intelligence with scholastic achievement was about .40. These authors point to adequate motivation, determination, and drive as possible factors related to school achievement.

PRESENT STUDY

The purpose of the present study was to investigate the following problems:

(1) How often is it possible for a student having an IQ below the mean of his class to achieve above the mean for his class?

(2) Do certain personality ratings improve the prediction of scholastic achievement? If so, what are these traits?

(3) What are the factors being measured by the variables used in this study?

Three graduating classes (1939, 1940, and 1941) of a mid-western high school were used as samples in the study of Problem One. Two hundred thirty students from the 1941 class were used as a sample in studying Problems Two and Three.

PROBLEM ONE

In this part of the study, we wished to know how many students in each of our three samples actually achieved grades above the

mean grade of their class even though their individual IQ's were below the average IQ for their class.

The intelligence of each of the students in all three classes was determined by the score made on an Otis S.-A. Intelligence Test administered by the school at the 9*B*-grade level. The index of achievement for each student was defined as the average grade earned during the four years of high school. Table I gives the comparison of the data from the three classes.

It will be noted that the mean IQ's, as well as the ranges in IQ's of these three classes are comparable to those reported in other studies. Some other interesting findings from the data of the 1941 class are that of the two hundred eighty-one students studied:

The IQ of the highest ranking student was.....	117
The IQ of the third highest student was.....	107
The IQ of the thirtieth ranking student was.....	91
Twenty-three students (eight per cent), who were in the upper one-half of their class in achievement, had IQ's of less than.....	100
The correlation of achievement with intelligence was [This correlation is comparable to those reported in other studies.]	.54

In this portion of our study we found that fifteen per cent of one class and between twelve and fourteen per cent of two other classes were able to achieve grades above the mean grade of their class even though their individual IQ's were below the average IQ for their class, all of which seems to indicate that factors other than intelligence are operating in the prediction of academic success.

PROBLEM TWO

In the second part of this study, an attempt was made to isolate some of the personality traits which might be supplementary to an IQ in the prediction of scholastic achievement, and in order to do this a personality rating scale was created by selecting twelve items from "A 36 Trait Personality Rating Scale,"[7] and adding a thirteenth item—general attitude toward school.

Each student was evaluated on this scale by each teacher having the student in class at the time of this study.

TABLE I.—COMPARISON OF DATA FROM THE THREE CLASSES

	YEAR		
	1939	1940	1941
Number of students...	256	277	281
Range of IQ's.....	68–137	78–128	76–132
Mean IQ.....	105 ± .6	104 ± .6	105 ± .7*
Mean achievement, 100 per cent = perfect..	84.06 ± .8	85.02 ± .9	85.08 ± .8*
Correlation between intelligence and achievement (*r*).....	.59 ± .04	.59 ± .04	.54 ± .05†
Number of students with IQ's below the mean of the class, but with achievement above the mean of the class.....	30	38	43
Same in percentage....	12%	14%	15%
Achievement of student with lowest IQ.....	77	81	79

* These are standard errors of the mean.

† Since there was little difference in the age of the students of the 1939 and 1940 classes, IQ's were correlated with achievement. The 1941 correlation is between achievement and the raw scores on the Otis S-A Intelligence Test.

Ratings were collected thus:

4 students were rated by five teachers.
146 students were rated by four teachers.
60 students were rated by three teachers.
20 students were rated by two teachers.

Each trait was rated on a graphic scale of 25 points, and the mean of the ratings for each item was taken as an index of that trait for each student. In order to determine whether a relationship existed among any of the fifteen variables measured, an intercorrelation table (II) was set up.

Three traits—sociability, attractiveness, and popularity—showed no significant correlation with scholastic achievement. Four traits—sociability, attractiveness, popularity, and courtesy—had no significant correlation with intelligence. However, persistence (.60 ± .05), common sense (.52 ± .05), and dependa-

bility (.57 ± .05) correlated with achievement about as highly as did intelligence (.54 ± .05). Punctuality (.47 ± .05), coöperation (.45 ± .06), honesty (.41 ± .06), and sincerity (.49 ± .03), also, had significant correlations with achievement. Since none of the thirteen traits correlated highly with the Otis score, it would seem that there are traits not measured by an Otis Intelligence Test.

The Doolittle method for computing the beta weights for each variable in the prediction of academic achievement was used, and the following beta weights established:

Dependent Variable—Achievement

1. Sociability	+.007	8. Persistence	+.365
2. Attractiveness	+.050	9. Honesty	+.031
3. Nervousness	+.010	10. Common Sense	+.055
4. Popularity	−.181	11. Sincerity	+.066
5. Punctuality	−.060	12. Dependability	−.047
6. Courtesy	−.048	13. Attitude	+.260
7. Coöperation	−.062	14. Otis Raw Score	+.353

The multiple *R* was found to be .72. Since the Otis score, attitude toward school, and persistence were the only beta weights of any size, new beta weights using these three variables alone were found. They were:

1. Otis raw score	+.358
2. Attitude	+.215
3. Persistence	+.303

The multiple *R* using only these three independent variables was .71 as compared to .72 using fourteen independent variables.

From this portion of our study we found that for predicting scholastic achievement from an Otis score alone the *r* was only .54, but for predicting scholastic achievement from an Otis plus ratings on persistence and attitude toward school the *R* is .71. Seemingly personality traits do correlate with scholastic achievement at the high-school level, at least, and much more accurate predictions of achievement can be made by using personality ratings plus an intelligence score.

It must be taken into account, of course, that the index of achievement used in this study was the average of all the grades received during the four years of high school. Some of these

TABLE II.—INTERCORRELATIONS*

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	..	.19	.15	.28	.14	.47	.32	.45	.60	.41	.52	.49	.51	.57	.54
1	..	...	.53	.21	.57	−.03	.17	.17	.36	.22	.35	.02	.22	.24	.18
2	..	...	...	.19	.65	.18	.36	.02	.28	.26	.32	.26	.22	.26	.11
3	..	...	...	...	.29	.41	.21	.39	.41	.36	.45	.40	.48	.38	.23
4	..	...	...	...	...	.28	.30	.23	.37	.39	.39	.33	.28	.33	.18
5	..	...	...	...	...	...	.56	.64	.69	.60	.60	.61	.75	.72	.39
6	..	...	...	...	...	...	...	.61	.54	.57	.55	.56	.59	.59	.18
7	..	...	...	...	...	...	...	...	.72	.62	.66	.69	.75	.76	.25
8	..	...	...	...	...	...	...	...	...	.66	.70	.68	.80	.77	.36
9	..	...	...	...	...	...	...	...	...	...	.68	.70	.72	.61	.27
10	..	...	...	...	...	...	...	...	...	...	...	.76	.73	.75	.37
11	..	...	...	...	...	...	...	...	...	...	...	...	.76	.76	.29
12	..	...	...	...	...	...	...	...	...	...	...	...	...	.91	.20
13	..	...	...	...	...	...	...	...	...	...	...	...	...	...	.34

* The variables are:

0. Four year average achievement.
1. How does he meet new social situations?
2. Is he physically attractive?
3. Is he nervous and "flies off the handle" easily?
4. Is he popular with other people?
5. How does he meet his appointments?
6. How courteous is he?
7. How coöperative is he?
8. How well does he stick to a task?
9. How honest and fair is he?
10. Does he possess common sense?
11. How sincere is he?
12. Can you count on him to do a thing?
13. What is the student's attitude toward school in general?
14. Otis Intelligence raw score.

grades were given by the same teachers who made the ratings on the rating scale. Therefore, while the grades may have been influenced by the personality of the student, the ratings made by the teachers may, in turn, have been influenced by the grades achieved by the student in classes with these teachers during the past year.

PROBLEM THREE

Using the Thurstone Centroid Method of Multiple Factor Analysis as described by Peters and Van Voorhiis,[8] the fifteen variables used in this study were factor analyzed. The results showed that two factors were measured.

Factor I was most effectively measured by variables 1, 2, and 4. These were:

(1) How does he meet new social situations?
(2) Is he physically attractive?
(4) Is he popular with other people?

Factor I seems to be a measure of the student's ability to succeed socially with his classmates. It will be recalled that these same items had no significant correlations either with intelligence or with achievement. The beta weights for variables 1 and 2 were practically zero, and the beta weight of variable 4 was −.18.

We conclude that the ability to succeed socially has no relation to high-school achievement.

Factor II was best measured by variables 12, 13, 7, 11, 5, and 8. These were:

(12) Can you count on him to do a thing?
(13) What is the student's attitude toward school?
(7) How coöperative is he?
(11) How sincere is he?
(5) How does he meet his appointments?
(8) How well does he stick to a task?

Factor II seems in this instance to measure the student's tendency to conform to the school situation. These items are highly correlated with achievement. The beta weights of persistence (+.365) and attitude toward school (+.260) carry most of the weight when predicting achievement.

From this part of our study, we find that our fifteen variables measure two factors. One is the student's ability to succeed socially. The other is the student's ability to conform to school situations. The first has no relation to scholastic achievement. The last is related to scholastic achievement.

SUMMARY

(1) The correlation between scholastic achievement and the score on an Otis Intelligence Test was .54 ± .05.

(2) The multiple *R* using the Otis score plus the thirteen-trait-rating-scale was .72.

(3) The multiple *R* for predicting achievement from the Otis raw score plus a two trait rating scale (persistence and attitude toward school) was .71.

(4) Two factors were being measured by the fifteen variables. These were the ability to succeed socially and the ability to conform to school situations.

(5) The ability to succeed socially was not related to scholastic achievement.

(6) The ability to conform to school situations was related to scholastic achievement.

BIBLIOGRAPHY

1. R. Byrons and V. A. C. Henmon: "Long Range Prediction of College Achievement," *School and Society*, Vol. XII, June 29, 1935, pp. 877–880.
2. H. A. Kohn: "Achievement and Intelligence Examinations Correlated with Each Other and with Teacher's Rankings," *Journal of Genetic Psychology*, Vol. LII, June, 1938, pp. 433–437.
3. F. J. Adams: "College Degrees and Elementary-school Intelligence Quotients," *Journal of Educational Psychology*, Vol. XXXI, 1940, pp. 360–368.
4. Noel Keys: "The Value of Group Test Intelligence Quotients for Prediction of Progress beyond High School," *Journal of Educational Psychology*, Vol. XXXI, p. 81.
5. V. E. Benson: "The Intelligence and Later Scholastic Success of Sixth-grade Pupils," *School and Society*, Feb. 7, 1942.
6. L. D. Hartson, and A. J. Sprow: "The Value of Intelligence Quotients Obtained in Secondary Schools for Predicting College Scholarship," *Educational and Psychological Measurement*, Oct., 1941, pp. 387–398.
7. E. Lowell Kelly: "A 36-Trait Personality Rating Scale," *The Journal of Psychology*, Vol. IX, 1940, pp. 97–102.
8. C. C. Peters, and W. R. Van Voorhis: *Statistical Procedures and Their Mathematical Bases*, McGraw-Hill Book Company, Inc., New York, 1940.

SOME RESULTS OF RETESTING ELEMENTARY-SCHOOL PUPILS WITH THE STANFORD REVISION OF THE BINET-SIMON TEST

LEO R. MILLER

Principal, Graceland School, Kansas City, Missouri

This study had its inception in an experience of the writer a few years ago, when he was working as an elementary-school teacher in a large Kansas City school. A girl in the fifth-grade group assigned to him attracted his attention. She was apparently quite dull, yet recent physical data showed that she had better than normal vision and hearing. Her IQ, based on a kindergarten test, was 122. In spite of these favorable factors, she was unable to cope with the fifth-grade material and could not keep pace with the other members of the class, an accelerated group.

The teacher went to the principal and asked for a retest on the girl. The principal considered this unnecessary, stating that there would be so little difference in the original and the retest IQ's that the retest would be a waste of time. However, upon the teacher's insistence, the retest was made. The retest IQ was 87. The girl, upon graduation from the elementary school, was placed in a special group in high school in which she could achieve some degree of success. Had the retest not been made and had she gone on to high school with the record of a 122 IQ, she would have been placed in an advanced group in which, the odds were, she would have met with a series of discouraging failures.

As a result of this experience, and others quite similar, the author has made it the general practice, since becoming a principal, to have each pupil tested with the Stanford Revision of the Binet-Simon individual intelligence test when he enters kindergarten or first grade and before he leaves the school to enter high school. As a result of this testing procedure, considerable data have accumulated bearing upon many phases of psychometrics.

Table I contains a summary of a segment of the amassed data. They deal with the results of testing ninety elementary pupils upon their entrance to grammar school and retesting them just before promotion to high school. The writer feels that they are

significant because of two characteristics: first, the relatively long period elapsing between tests and, second, the comparatively large average variation in the two resulting sets of IQ's.

TABLE I.—SHOWING AMOUNTS OF VARIATION IN SCORES OF NINETY ELEMENTARY-SCHOOL SUBJECTS BETWEEN ORIGINAL TEST AND RETEST

Points of variation.....	0	1	2	3	4	5	6	7	8	9
Number of pupils.....	2	9	8	1	5	3	5	3	5	8
Points of variation.....	10	11	12	13	14	15	16	17	19	20
Number of pupils.....	5	3	1	3	4	6	2	3	2	1
Points of variation.....	21	22	23	24	26	28	30	32	34	
Number of pupils.....	1	2	1	2	1	1	1	1	1	

These tests were, in both cases, administered by testers from the Psychological Clinic of the Kansas City public schools. These testers have had years of experience and have tested thousands of boys and girls of all grades and ages. In some instances the pupils reported in these data were tested on both occasions by the same tester, in the same building, and in the same room. While mistakes on their part would be inevitable, the writer feels that they are as competent as the testers furnishing any other group of intelligence test data.

There is good reason to believe that there are many teachers, and administrators, who, like the principal mentioned, do not see any necessity for retesting as a general policy. If the writer's memory is not amiss, there is a statement in the *Encyclopedia*

COMPARISON OF RESULTS OF THIS STUDY WITH SIMILAR DATA FROM TERMAN-MERRILL (P. 51)

	CA First Test (Years)	SD CA First Test	IQ First Test	SD IQ First Test	CA Second Test (Years)	SD CA Second Test	IQ Second Test	SD IQ Second Test	Mean Interval in Years between Tests	Mean of the Variations in IQ Points
Terman & Merrill.....	6	.....	98.0	13.25	11.42	....	94.6	15.95	5.42	
Present Study.....	5.86	19.98	104.8	28.8	12.30	25.4	108.4	33.2	6.5	10.36

of Educational Research under the topic heading, "The Constancy of the IQ," to the effect that, while IQ's fluctuate, the variation will average about five points.

The author feels that many of the studies upon which that assumption is based were "shotgun" studies, made to fit the schedules of candidates for advanced degrees. In the present study, however, the period between original tests and retests has not been telescoped to facilitate the completion of a dissertation. On the contrary, that period averages seventy-eight months or 6.5 years, spanning the entire grammar-school experience of the subjects. In every instance, the original test was the Stanford-Binet, 1916 edition and the retest was the 1937 revision, Form *L*.

These ninety subjects were not hand-picked. They are the ones most recently retested in the Graceland School. The retests cover a period of four years, the last ones being administered in the Spring of 1942.

Curiously, the ninety subjects divide equally between boys and girls. Of the fifty-three pupils showing plus variations in IQ, twenty-two were girls and thirty-one were boys. Of the thirty-five showing minus variations, twenty-three were girls and twelve were boys. Two boys showed exactly no variations over the period between the tests. The average of the plus variations alone was 11.8 points and of the minus variations alone, 8.5 points.

CONCLUSIONS

Boys are more likely than girls to show a gain in IQ points over a relatively long period between original test and retest. Conversely, girls are more likely to show a loss in IQ points over a similar period.

The administration of individual intelligence tests should not be primarily for the purpose of establishing norms for administrative or clinical purposes, but should be principally for the purpose of helping individual pupils make better adjustments in choices of courses and selections of vocations.

Relatively large shifts in IQ's are so common, even in a small population, that, without retests, severe injustices could be worked upon individuals under any plan of homogeneous grouping based upon IQ's.

In fact, the study on which this report is based furnishes ample evidence for believing that individual intelligence tests, administered in the kindergarten or first grades and used as a basis for ability grouping in later grades or high school without benefit of retesting, could do more harm to some pupils than could the failure to give any intelligence test at all.

A COMPARISON OF THE VALIDITY AND RELIABILITY OF THREE TYPES OF OBJECTIVE EXAMINATIONS

JOHN S. COPELAND AND A. R. GILLILAND

Northwestern University

Several studies have compared the merits of essay and objective examinations.[1] Even more study has been devoted to the relative merits of different types of objective examinations. For example, Ruch and Charles[2] compared five types of objective examinations and found that the true-false and multiple-response questions measure the same functions as do the completion forms. Eurich[3] says, "the essay, multiple-choice, completion, and true-false tests have approximately equal validity." Similar results were obtained by Magill[4] and Shulson and Crawford.[5] But Bayless and Bedell[6] found controlled completion tests to have the highest validity with modified true-false tests second, multiple-choice third, and matching fourth. Peters and Martz[7] state "the most valid is the multiple-choice test followed in order by the completion, essay, and true-false." Watson and Crawford[8] found completion tests most valid, followed in order by essay, best answer, and true-false. Carter and Crone[9] found

[1] Eurich, A. C.: "Four Types of Examinations Compared." *J. Ed. Psych.*, Vol. XXII, 1931, pp. 268–278; and Gilliland, A. R. and Misbach, L. E.: "Relative Values of Objective and Essay Type Examinations in General Psychology." *J. Ed. Psych.*, Vol. XXIV, 1933, pp. 349–361.

[2] Ruch, G. M. and Charles, J. W.: "A Comparison of Five Types of Objective Tests in General Psychology." *J. Appl. Psych.*, Vol. XII, 1928, pp. 398–403.

[3] Eurich, A. C.: *op. cit.* 270.

[4] Magill, W.: "The Influence of Form on Validity." *J. Ed. Psych.*, Vol. XXV, 1934, pp. 26.

[5] Shulson, V. and Crawford, C. C.: "An Experimental Comparison of True-False and Completion Tests." *J. Ed. Psych.*, Vol. XIX, 1928, pp. 580–583.

[6] Bayless, E. E. and Bedell, R. C.: "A Study of Comparitive Validity as Shown by a Group of Objective Tests." *J. Ed. Res.*, Vol. XXIII, 1931, pp. 8–16.

[7] Peters, C. C. and Martz: "A Study of the Validity of Various Types of Examinations." *School and Society*, Vol. XXXIII, 1931, pp. 336–338.

[8] Watson, D. R. and Crawford, C. C.: "Four Types of Tests." *High School Teacher*, Vol. VI, 1930, pp. 282–283.

[9] Carter, H. D. and Crone, A. P.: "The Reliability of New-type or Objec-

completion questions to have the highest reliability followed by multiple choice and true-false. This is in agreement with the results of Bird and Andrew.[1]

Because of the diversity of these results, the present study has been conducted in an attempt more rigorously to control the variables involved. This was accomplished by (1) using three groups of subjects of equal ability, (2) rotating the types of tests used, (3) using the same subject-matter as the basis for all tests, (4) using equal care in the construction of the various types of tests, and (5) equating the time spent in answering each type of test.

The merits of each type of examination were determined by their respective validities and reliabilities.

One hundred eighty members of a class in child psychology who took all the examinations were divided into three groups of sixty each. These were called groups *A*, *B*, and *C*. The mean grades of each group for the preceding semester were 3.52, 3.73, and 3.89, respectively. By means of the Chi-square technique it can be shown that differences as great as this may occur in ninety-nine cases out of one hundred as a result of chance. The mean scores in mental alertness tests for the three groups were as follows: 156.24, 148.66, and 153.74, respectively. Differences as great as this may occur in ninety cases out of one hundred as a result of chance. In terms of final grades in the course in child psychology, differences as great as those found between the three groups would result by chance in ninety-five cases out of one hundred. Therefore, it is safe to assume that the three groups were of equal ability.

The three types of tests used were true-false, multiple-choice (four choices), and matching (each answer used once only). There were twenty questions in each examination. During the first three tests each member of each group took one of each type of examination; that is, the type of examination was rotated for each group. This was followed by a fourth examination in which

---

tive Tests in a Normal Classroom Situation." *J. Appl. Psych.*, Vol. XXIV, 1940, pp. 353–368.

[1] Bird, C. and Andrew, D. M.: "The Comparative Validity of New-type Examinations." *J. Ed. Psych.*, Vol. XXVIII, 1937, pp. 241–258, and "A Comparison of Two New-type Questions: Recall and Recognition." *J. Ed. Psych.*, Vol. XXIX, 1938, pp. 175–193.

each person took the same examination composed of six true-false questions, seven multiple-choice questions, and seven matching questions. For the next five examinations group *A* took the true-false type; group *B*, multiple-choice; and group *C*, matching questions. The last test in the series was like the fourth examination mentioned above—that is, all the students took the same examination which was composed of six true-false, seven multiple-choice, and seven matching questions.

The questions were all based in order upon Chapters II to XI of Morgan's text, *Child Psychology* (Revised Edition). As already mentioned, each examination was composed of twenty questions. Each list after being constructed was checked by two other psychologists in order to insure maximum clarity of statement and as nearly as possible equally good questions of each type. No method seemed available for measuring equal quality in the questions of each type, but it seems probable that there was no significant difference between them.

The mean time required for completing each type of examination was determined by having the students note the time when they completed their tests. It was found that the true-false required the shortest time, the multiple-choice was second, and matching required the most time. The relative mean times for the three types were as follows: 18.7, 26.8, and 54.5, respectively.

There was also a difference in the difficulty of the different kinds of questions. Since the questions were based upon the same subject-matter, it seems reasonable to assume that the following percentages of errors at least roughly represent the varying difficulty of the types of tests: true-false, twenty-five per cent; multiple-choice, thirty-one per cent; and matching, forty-four per cent.

In comparing different types of examinations it is probably fairer to use tests that require the same length of time to take, rather than tests made up of the same number of items. Therefore, the different types of tests were equated for time by means of the Spearman-Brown formula.

The validity of the three kinds of tests was calculated in two different ways. The mean scores based on six tests of each type were correlated with the intelligence test scores of the students. The correlations are as follows: true-false .07, multiple-choice .26,

and matching .25. The other method of validating was by a comparison of the mean grades with the point averages of these students for the preceding semester. The correlations were as follows: true-false .35, multiple-choice .20, and matching .57. As can be seen from these results the validities as measured by these methods are not high and they are especially low when intelligence test scores are used. While there may be some question about whether this is an entirely satisfactory method of measuring validity, at least it gives a comparison of the relative validity of the different types of examination.

In order to determine the reliability of the three types of examinations, mean grades for tests 1, 5, and 6 and for tests 7, 8, and 9 were determined. The reliability of the three types are shown in Table I.

TABLE I.—RELIABILITY OF THREE TYPES OF OBJECTIVE TESTS

	TRUE-FALSE	MULTIPLE-CHOICE	MATCHING
Three tests with three tests (calculated)	.38	.39	.62
One test with 1 test (predicted)	.17	.18	.36
Six tests with 6 tests (predicted)	.55	.56	.77

These correlations show a considerably greater reliability for matching questions than either of the other kinds. But, as previously pointed out, students spend almost twice as much time in answering matching questions as multiple-choice questions and considerably less time for true-false questions than either of the other two forms. When equated for time in taking the tests using matching questions as a basis for comparison, the reliabilities become, true-false .62, multiple-choice .54, and matching .62, when six tests are compared. Therefore, when enough questions are added to make the three types of examinations equal in length of time necessary in answering the questions, there is no difference between true-false and matching tests, and multiple choice tests are only slightly less reliable (c.r. = .45).

SUMMARY AND CONCLUSIONS

(1) The relative amount of time required to complete examinations consisting of twenty questions each were: true-false, 18.7; multiple-choice, 26.8; and matching, 54.5.

(2) The relative difficulty of the different types were: true-false, .25; multiple-choice, .31; and matching .44.

(3) The validities of the different types of tests when correlated with intelligence test scores were: true-false, .07; multiple-choice, .26; and matching, .25. When correlated with previous semester grades they were: true-false, .35; multiple-choice, .20; and matching, .57.

(4) Reliabilities as obtained by correlating means of six tests with six other tests were: true-false, .55; multiple-choice, .56; and matching, .62.

(5) When the tests were equated for the time required for taking the tests, these reliabilities become: true-false, .62; multiple-choice, .54; and matching, .62.

This study indicates that when other variables are carefully controlled, matching questions are somewhat more valid than either true-false or multiple-choice questions. When the time necessary for taking examinations is equated, there is no difference between the reliability of true-false questions and matching questions, while multiple-choice questions seem to have a slightly lower reliability. But these differences are so small that in actual practice it makes relatively little difference which type of question is used if the examinations are of equivalent length in terms of time necessary for taking the test.

ILLUMINATION INTENSITIES FOR READING NEWSPAPER TYPE*

MILES A. TINKER

University of Minnesota

From the viewpoint of hygienic vision, the critical level of illumination intensity may be defined as the brightness below which there is a decrease in efficiency of performance. The performance may consist of type-setting, sorting mail, reading, machine operations, etc. It has been found that the critical level of light intensity for reading ten-point book type is about 3 foot-candles.[1,2] For intensities less than 3 foot-candles, speed of reading decreased significantly. When the light intensities were varied from 3 to 50 foot-candles, however, there were no changes in rate of reading.

Observation of ordinary newspaper type suggests that reading the newspaper is a rather severe visual task. Most newspapers are printed in relatively small type (seven or eight point) on a poor quality of paper. It is possible, therefore, that the critical level of illumination intensity for reading newspaper type is different from that for reading ordinary book type.

The purpose of this study is to determine the critical level of light intensity for reading newspaper type. The text to be read was printed in regular newspaper typography. The light intensities employed ranged from a level that was obviously too dim for clear seeing to one that was considerably above the intensities commonly employed for reading newspapers. The seven intensities used were 1, 4, 7, 10, 20, 50, and 100 foot-candles. A recently calibrated General Electric light meter was employed to measure the intensities which were rechecked from day to day.

The experiment was conducted in a special light laboratory, fifteen by nine by eleven and five-tenths feet. Indirect illumi-

* Grateful acknowledgment is given to the Graduate School, University of Minnesota for research grant to finance this study. The writer also gratefully acknowledges the aid of Mr. H. I. Kennedy and the Pittsburg Reflector Company for furnishing the equipment for the lighting laboratory; to Mr. P. T. Owens of the Northern States Power Company who helped plan the installation of the equipment, and to Dr. W. F. Holman, University of Minnesota, who furnished facilities for installing the equipment.

nation was derived from five Pittsburg Permaflector luminaires suspended twenty-seven inches from the ceiling. The reading table was approximately in the center of the room.

The reading material consisted of Forms *A* and *B* of the Chapman-Cook Speed of Reading Test. Performance on Form *B* is equivalent to that on Form *A* on the average. A control group was introduced to check on this equivalence.[3] Both forms of the test were printed in Ionic No. 5 newspaper type, seven point, set solid, twelve pica line width on newsprint paper stock.

Seven groups of university sophomores served as subjects, four hundred five in all. The number in each group ranged from fifty-three to seventy. The subjects were measured one at a time. Each subject read Form *A* of the reading test as a standard under 10 foot-candles of light. Form *B* was then read as follows: Under 1 foot-candle in Group I, 4 foot-candles in Group II, 7 foot-candles in Group III, 10 foot-candles in Group IV (control group), 20 foot-candles in Group V, 50 foot-candles in Group VI, and 100 foot-candles in Group VII. The order of testing was systematically varied. The use of this technique makes it possible to determine the speed of reading under each level of illumination in comparison with the standard. The data will then reveal the brightness level beyond which there is no further improvement in performance with increase in light intensity.

In an earlier investigation[1] it was shown that the condition of visual adaptation is important in studies of this kind. Each subject, therefore, was adapted for fifteen minutes to the illumination under which the reading was to be done.

RESULTS

The results are given in Table I. Note that in Group IV (the control group) the mean scores for Forms *A* and *B* are practically identical. This indicates equivalence of the two forms when experimental conditions are constant. In Group I, .98 fewer paragraphs were read under 1 foot-candle than under 10 foot-candles. This is a highly significant difference. Similarly, the text was read significantly slower under 4 foot-candles than under 10 foot-candles. In this study, a significant difference is defined as one which is significant at the one per cent level.

TABLE I.—EFFECT OF LIGHT INTENSITIES UPON SPEED OF READING NEWSPRINT

Test Group	N	Foot-candles for B	Mean Form A	Mean Form B	Difference A-B	D.F.	Value of t	Level of Significance
I	61	1	19.16	18.18	−.98	60	2.78	1
II	55	4	18.96	18.11	−.85	54	2.80	1
III	70	7	18.21	18.14	−.07	69	0.09	90
IV	55	10	20.16	20.21	+.05	54	0.14	90
V	53	20	19.50	18.80	−.70	52	2.04	5
VI	55	50	18.60	18.36	−.24	54	0.72	50
VII	56	100	19.80	19.67	−.13	55	1.05	30

NOTE.—Differences given are for mean score on Form *A* minus mean score on Form *B*, Chapman-Cook Speed of Reading Test printed in Ionic No. 5 newspaper type, seven-point set solid, twelve pica line width on newsprint paper stock. Form *A*, the standard, was read under 10 foot-candles of light. N = 405 university students.

When the 7 foot-candle level of light intensity is reached, the trend of the findings changes. For 7, 10, 20, 50 and 100 foot-candles, the differences between the standard and the variables are not statistically significant. In other words, rate of reading does not change significantly when the illumination is varied from 7 to 10, 20, 50 and 100 foot-candles. It will be noted, however, that the difference between reading under 20 foot-candles and 10 foot-candles is .70 of a paragraph. This is larger than the difference found for 7, 10, 50 and 100 foot-candles. Since this difference for 20 foot-candles is not large enough to be statistically significant, the trend of the results from 7 to 100 foot-candles remains constant.

The pattern of the data in Table I indicates that the critical level of illumination intensity for reading newspaper type is at approximately 7 foot-candles. Below 7 foot-candles speed of reading is retarded. At 7 foot-candles and above (to 100 foot-candles), the rate of reading is not significantly influenced by changes in light intensity. This is in contrast to the critical level of 3 to 4 foot-candles for reading book type.[2]

As pointed out in a prior publication,[2] reading should not be done at the critical level of illumination. There should be a margin of safety, *i.e.*, a somewhat brighter light should be used. In the case of newspaper type (seven point), therefore, the level of illumination should probably be 15 to 20 foot-candles.

SUMMARY

(1) This study was planned to determine the critical level of light intensity for reading newspaper type. The critical level is defined as the intensity beyond which no further change in reading performance occurs as the intensity is increased.

(2) Text printed in newspaper type was read under 1, 4, 7, 10, 20, 50 and 100 foot-candles.

(3) Seven groups of university students (four hundred five in all) were tested, one subject at a time.

(4) The critical level for reading seven point, newspaper type, set solid, was found to be approximately 7 foot-candles.

(5) To provide a margin of safety, it is suggested that 15 to 20 foot-candles of light be employed when reading newspapers.

REFERENCES

(1) M. A. Tinker: "The effect of illumination intensities upon speed of perception and upon fatigue in reading." *J. Educ. Psychol.*, Vol. xxx, 1939, pp. 561–571.

(2) M. A. Tinker: "Illumination standards for effective and comfortable vision." *J. Consult. Psychol.*, Vol. III, 1939, pp. 11–20.

(3) For detailed description of methodology and controls see Tinker, M. A. and Paterson, D. G.: "Studies of typographical factors influencing speed of reading. XIII. Methodological considerations." *J. Appl., Psychol.*, Vol. xx, 1936, pp. 132–145.

BOOK REVIEWS

KATHERN M. MCKINNON. *Consistency and Change in Behavior Manifestations.* New York: Teachers College, Columbia University, Child Development Monographs, No. 30, 1942, pp. 144.

This report is concerned with evidences of change and persistence in children's behavior and personality characteristics over a five- or six-year period together with the effect of teachers' guidance procedures on change and persistence in the behavior manifestations. Eight boys and eight girls were observed from time of entrance to nursery school until about the eight-year year level. Data were derived from accumulated school records and direct observation. It was possible to characterize the most conspicuous quality of behavior of each child. The behavior of five children was predominantly characterized by 'conformity,' five by 'invasiveness,' three by 'withdrawal,' and three by 'caution.' For ten children, the dominant personality characteristics manifested at three years continued throughout the period of observation (five or six years). The least likely to change was withdrawing behavior. For those whose behavior did change, the tendency was in the direction of conformity. Predominant conformity, however, tended to be associated with characteristics which limited personal development in four of the seven children in this classification. Also it seemed that predominantly invasive, cautious and withdrawing forms of behavior were detrimental to well-rounded personality development. In some cases, improvement in skills appeared to have a beneficial influence on adjustments. Implications for guidance are given.

One may raise certain questions concerning methods and conclusions in this study: (*a*) The observational situation was an experimental school and the population studied appears to be non-typical of school children in general. Application of suggestions on guidance, therefore, must of necessity be limited. (*b*) All four of the listed predominant manifestations of behavior appeared in general to be detrimental to personal adjustment. This raises an unanswered question concerning the direction that guidance is to take. The author places little emphasis upon development of a balance between these behavior characteristics. (*c*) Since the prevailing trend in change of behavior manifesta-

tions was toward conformity, the reader may wonder if this would not have occurred without guidance. No control group was studied. (*d*) The meaning of the data in many instances is not unequivocal. For instance, many inferences are introduced by such phrases as "it is possible," and "it is probable."

This investigation represents an interesting technique of child study. Valid inferences for guidance, however, must await an extension of the method with certain modifications to more representative populations.

MILES A. TINKER.
University of Minnesota.

LAURA W. DRUMMOND. *Youth and Instruction in Marriage and Family Living.* New York: Bureau of Publications, Teachers College, Columbia University, 1942, pp. 186.

The phases of marriage and family life that young people are most interested in are sex adjustments; premarriage problems, including dating, courtship, and choice of mate; accord in marriage and family relationships; and family economics. Problems in which they are moderately interested are child-parent relationships and adjustments between generations. Problems in which they are least interested are divorce problems, religion, and the family as an institution.

These are the general conclusions reported by the author of this volume on *Youth and Instruction in Marriage and Family Living.* The author interprets these conclusions to imply that young adults are more concerned with the kind of persons they ought to marry than about the money needed to marry with; more concerned with specific situations than with generalizations; with personal problems than with social theories. Her findings are based on a tabulation of 2,752 suggestions concerning instruction in marriage and family living received from three hundred twenty freshmen, one hundred sixty-four seniors, and one hundred forty-eight alumni from two large Pennsylvania universities. The reliability of the information is presented in analyzed form by the use of the Chi-Square test of significance and analysis of variance. A summary of the findings is presented in the first chapter. For the type of information reported, this is appropriate. A discussion of the questions is presented in some detail.

Information presented as well as methods and implications considered should be of value to persons who give courses in family

counselling on college campuses. Clinicians will probably find lacking here the type of information that can be obtained only by interviewing rather than by this type of response to a free questionnaire. A supplementary investigation of some of these people by the use of clinical interviewing might have revealed more dynamic facts that are often of greater significance in actual human relations and particularly family living. H. MELTZER.

Psychological Service Center, St. Louis, Missouri.

WALTER L. DEEMER AND PHILLIP J. RULON. *An Experimental Comparison of Two Shorthand Systems.* Cambridge, Mass.: Harvard University Press, 1942, pp. 294.

The work of these authors stands as an empirical rebuttal to the skeptical psychologists who maintain that experiments in education do not demonstrate anything conclusively. The authors set out to determine "the relative merits of Gregg and Script shorthand when taught in public high schools for a period of two years, with the objective of general office use." By controlling the teaching and its supervision, by matching the pupils studying under each system on pertinent variables, and by having tests administered by trained examiners under carefully controlled conditions and scored by a central office under close supervision, the authors could conclude beyond any doubt that "in terms of the features examined in the study, Script shorthand exhibited substantially greater relative merits than Gregg."

The results are presented in the form of twenty-six comparisons between the Gregg and Script shorthand systems. In twelve of these twenty-six comparisons, the Script students excelled the Gregg, and, according to the authors, the differences are statistically significant. In two of the comparisons the superiority is in favor of the Gregg system, the difference again being statistically significant. In the twelve remaining comparisons, nine favored the Script students and three the Gregg students, but none of these differences were large enough to permit their being labelled 'statistically significant.' The number of cases on which the comparisons were based varied from three hundred seventy-five to seven hundred fifty-three.

The method of analyzing the comparative data is an extension of the Johnson-Neyman test of linear hypotheses. The nature of the analysis is such that group results for the matched groups

are obtained and that "if method *A* is better than method *B* for bright pupils but *B* is better than *A* for dull pupils, this fact will be shown." Further, this method of analysis has the advantage that it is not necessary to drop pupils from one group or the other in order that the mean scores of the two groups on the matching variable be equal. An exact description of the method is given in detail in a longer and more technical report by Walter L. Deemer, *An Empirical Study of Relative Merits of Gregg Shorthand and Script Shorthand,* to be published by Harvard University Press. The fourteen comparisons where the differences were statistically significant are shown on a two-dimensional plot of the individuals, the *axes* being the variables on which the pupils were matched. Thus, one can observe for the entire range of the matched variables where the differences are significant. On the charts used for this purpose, lines were plotted showing a region of highest significance and a region of high significance. The region of high significance is that bounded by a line marked '$P = .05$,' indicating that the chances are only five in one hundred that the difference between Script and Gregg would have been obtained if there really had not been any difference in the two systems. Similarly, a line marked '$P = .01$' bounds the region of highest significance. For the '$P = .01$' group, the chances are only one in a hundred that the difference might have occurred when no real difference existed.

The results are given for four groups of paired students, two groups of non-typists who wrote transcripts in long hand, and two groups of typists who typed transcripts. Results are reported for these groups in terms of accuracy in transcribing notes taken at eighty words per minute or less and speed of transcription of these notes, in terms of accuracy in transcribing notes taken at ninety words per minute or more and speed of transcription of these notes, in terms of accuracy and speed in transcribing notes ('cold' notes) two weeks after they were taken, and in terms of accuracy in transcribing new words not appearing in the text books used by the students. For one group of paired students the Gregg students excelled the Script significantly only in accuracy of transcription of fast dictation (ninety words per minute or more) and in speed of transcription from 'cold' notes. However, for the latter variable there was conflicting evidence in that for another group of paired students the Script students excelled the Gregg significantly. For all remaining variables in all groups, the Script students either

excelled the Gregg students significantly or the differences between the groups had no statistical significance.

The authors cite incidental findings as to achievement which indicate that in some cases a "certain unusual type of student" may do better with Gregg than with Script.

A follow-up of some of the students was made by means of questionnaires to the students and visits to their employers. The number of cases upon which the results of the follow-up was based is too small to warrant any conclusions to be drawn from them. Most of the differences between Script and Gregg are reported by the authors as not having statistical significance.

The study would have been even more conclusive had the results of the follow-up been more complete. Also, as the authors themselves state, "a truly complete study of the relative merits of shorthand systems should make all possible comparisons (in colleges, business schools, high schools; between personal use, general office use, and court and convention reporting)." The data required for these comparisons were beyond the resources of the study so that it had to be restricted to one comparison; namely, that between shorthand users in high schools.

MOREY J. WANTMAN.
University of Rochester.

GEORGE K. BENNETT and RUTH M. CRUIKSHANK. *A Summary of Manual and Mechanical Ability Tests.* (Preliminary Form). New York: Psychological Corporation, 1942, pp. 75.

School psychologists will find this compilation of information about tests of manual and mechanical abilities very helpful. It is important, however, always to sound the caution that most of the tests reported in this review are more useful for vocational selection than vocational guidance purposes. Indeed, when we remember that it is the probable ability of a person which we desire to forecast in vocational guidance, we realize that an accurate measure of present abilities has serious limitations. This may be especially true in manual dexterity which is subject to psychological parameters which may be considerably different in the adolescent boy or girl than in the adult man or woman.

The tests are classified as follows: (1) Mechanical ability—paper and pencil tests, (2) Mechanical ability—performance tests

with apparatus, (3) Manual ability—mainly performance tests with apparatus. A general description is given in the Introduction, *viz.*:

"An individual description of a larger number of these tests, standardized to some extent for American populations, is given in a later section. . . . These test summaries include such pertinent information regarding each test and its applications as would be of help in determining whether a test should be examined more thoroughly for use. There are probably some tests summarized which are of limited value, and there are probably a number of other tests listed by titles which, upon further experimental use, may prove of greater intrinsic worth than others now available."

Each test treated in full is clearly and concisely set out, under complete reference headings—author, publisher, age range, date, approximate price, scoring methods, testing and scoring times, type of test, and test content. Standardization is given under: Reliability, Validity, and Application. A list of articles or references, under the heading Reviews and Research, is given for each test.

One section deals with the prediction of vocational success as related to tests of mechanical and manual ability. Prediction of success in the fields of engineering and physics, and of dentistry, medicine and surgery, is discussed, with some tables of correlations in the use of tests. A short section relates to the use of manual and mechanical tests in vocational selection.

Another section of the review is devoted to problems in the measurement of mechanical and manual ability, in which the effect of practice upon performance, determination of validity, establishment of a critical score, and sex differences are noted. Following the test summaries, the survey is completed by a list of annotated titles of other tests which, for various reasons, it is not thought necessary to summarize fully in the present publication.

A table of general references, other than the bibliography attached to each test summary, and an alphabetical index of all tests referred to are also included. This publication will be helpful to teachers and administrators who are counselling students, and to "research workers interested in the analysis of human abilities."

FORREST H. KIRKPATRICK.

Bethany College, W. Va.